Human Capital and Migration: a Cautionary Tale*

Salvador Navarro† and Jin Zhou‡

†University of Western Ontario
‡University of Chicago

July 15, 2020

Abstract

We analyze the impact that the option of migration might have on human capital accumulation. As long as the return to education for individuals is lower in the destination than in the origin, this reduces the overall incentive to accumulate human capital, compared to a situation in which migration is harder or impossible. We show, both with quasi-experimental and structural methods, that this indeed happened in China after the 1983 reform that eliminated the strong restrictions that existed for rural-urban migration. Since the return to education for rural migrants in the city is lower, the reform resulted in a reduction in average years of schooling of almost half a year for rural people in China.

*We thank Dan Ackerberg, Steven Durlauf, James Heckman, Terry Sicular, Jeff Smith, Chris Taber and Sergio Urzua for helpful discussions. We acknowledge support from the Social Sciences and Humanities Research Council of Canada. Salvador Navarro is at the University of Western Ontario, E-mail: snavarr@uwo.ca. Jin Zhou is at the Center for the Economics of Human Development, University of Chicago, E-mail: jinzhou@uchicago.edu

1 Introduction

Throughout his career, James Heckman has stressed the importance of being problem-oriented as opposed to methods-oriented when studying economic problems. As such, he has created, extended, or employed whichever approach was better suited to the problem at hand. Some of the many examples one could cite include: his development of selection corrections (Heckman, 1974) to study female labor supply; the use of factor models to generate instruments to study bundling (Heckman and Scheinkman, 1987); the use and development of diff-in-diff matching methods to study job training programs (Heckman et al., 1997); using general equilibrium models to study tuition policies (Heckman et al., 1999); the development of local instrumental variables to study marginal and policy relevant treatment effects (Heckman and Vytlacil, 1999); the use of dynamic factor models to study the technology of skill formation (Cunha et al., 2010a), etc. This paper continues the Heckman tradition by studying the effects that immigration policy can have on the potential immigrants accumulation of human capital using both a difference-in-differences approach and a dynamic Roy model of immigration (Roy, 1951; Heckman and Honore 1990) as our tools.

The study of how the characteristics of immigrants differ from those of non-immigrants and natives has a long history in Economics (e.g., Chiswick, 1978; Borjas, 1987). Whether immigrants are positively or negatively selected in terms of human capital/skills determines their effects on the wage distribution, both in the sending and receiving locations (Borjas, 2003). As a consequence, the literature has mostly taken the relative distribution of human capital/skills of migrants as given when studying the impact of migrants at home and/or abroad (e.g., Borjas, 1995; Chiquiar and Hanson, 2005 or Card, 2005). In this paper, we study the effect that the option of migration has on labor outcomes of migrants. However, we do not take the human capital of the individual as given. Instead, we allow for the possibility that, while human capital determines migration, the mere existence of the migration option can also affect the individual's human capital accumulation decision.

The point of departure of our analysis is a joint Roy model of migration and education decisions that leads to the following simple observation: if the return to human capital accumulation for migrants is lower in the potential receiving location than it is at home, then the overall return to human capital is lower than what it would be if the possibility of migration did not exist.[1] This pattern of returns is likely to exist in places in which low skilled migration is prevalent, and where the types of jobs available for these migrants in the receiving location are not skill intensive. For example, if a high school dropout and a high school graduate migrant will both end up washing dishes at a restaurant, then the skill component of the return to migration is likely to be very low. When this is the case, the mere possibility that the option of migration is there, lowers the incentive to accumulate human capital in the source location, when compared to a case in which

[1] See Black et al. (2005) for a related point regarding changes in returns.

migration is not possible (or harder).

Our analysis is in contrast to the standard assumption that the return to education (human capital) is higher at the migration destination. In this case the migration option makes individuals in the less developed source locations have more incentive to invest in their human capital (e.g., Dustmann and Glitz, 2011). However, as shown by Hendricks and Schoellman (2017), when high skill and low skill workers do not perfectly substitute, this needs not be the case. Furthermore, the return mechanism we highlight is different than the one studied in the brain drain literature (Maria and Stryszowski, 2009; Maria and Lazarova, 2012) which argues that migration will change the composition of skills due to the emigration of skilled people. It is closer to the idea motivating the work of Stark et al. (1997), Mountford (1997), Vidal (1998), Beine et al. (2001), Stark and Wang (2002), Beine et al. (2008) where the possibility of migration raises the return to education and may lead to a higher level of human capital.

In order to examine the empirical plausibility of our hypothesis, we use the Chinese Household Income Project panel to analyze the human capital accumulation effects of one of the largest migration episodes in human history: the rural to urban migration in China that started in the mid 1980's. The literature on rural education in China has, for the most part, ignored the impacts of the migration option and has focused on other determinants like poverty (Knight et al., 2009), or the ineffectiveness of the college expansion (Li et al., 2015). Brauw and Giles (2017) are the closest to our study, in that they empirically find that the reduction of migration costs has had a negative relationship with high school enrollment. They, however, do not propose a mechanism that could explain this finding.

The case of China is a good fit for our analysis. As a consequence of the segregation policy of 1958, people in the rural areas were prohibited from migrating to urban areas in China until the early 1980's. At this time, the Chinese government began implementing migration policies to first allow and then gradually facilitate and encourage more and more rural people to migrate to urban areas. We first show that, in a simple regression comparison, the returns to education between rural and urban areas are consistent with what is required in our simple model. We then exploit the differential educational attainment between rural and urban areas before and after the implementation of the policy relaxing the migration restrictions. In a difference-in-difference (DiD) analysis, we show that the relaxation of the migration restrictions was negatively associated with years of education for people born in rural areas.

The evidence from the difference-in-differences estimates is consistent with the story that our simple Roy model suggests. However, the extent to which the policy actually reduced migration costs (i.e., facilitated migration) is unknown. Other policies that could affect educational attainment were also implemented (and have been implemented since). Therefore, in order to quantify and distinguish the migration policy effect from other policies (e.g., the college expansion policy), we develop and estimate a life-cycle dynamic discrete choice model in which we try to account not only for the migration policies implemented in the early 80's, but also other policies like the college expansion policy, etc. In the model, individuals differ in terms of two types of unobserved skills, cognitive and non-cognitive skills (e.g., Heckman et al., 2006, Heckman and Mosso, 2014, Navarro and Zhou, 2017), and these characteristics can affect their educational attainment and location choices.

The model estimates show that the return to education in rural areas is higher than that in urban areas, consistent with the findings in our regression analysis. Second, we also find that the different types of unobserved skills play different roles. For example, cognitive and non-cognitive skills are rewarded differently in rural and urban areas. In urban areas, cognitive skills have a positive return, while non-cognitive skills have a positive return in rural areas. Individuals with high non-cognitive skills have lower migration costs.[2] We find that the implementation of the policy in the early 80's reduced migration costs significantly, and that this lead to a significant reduction in rural individuals human capital accumulation, consistent with the DiD evidence. In our counterfactual analysis, we eliminate the migration policies and find that the average educational attainment of rural individuals increases around one third of a year, an estimate that matches closely the one obtained in the DiD analysis.

The rest of the paper is organized as follows. Section 2 develops a simplified Roy model of migration and education that delivers the conditions under which having an option to migrate can reduce human capital accumulation in the source location. Section 3 presents background information on the education system and rural-urban migration in China, and describes the data in detail. Section 4 takes advantage of the migration policy change in a DiD framework to provide evidence consistent with the prediction of our simple model. In Section 5 we develop a dynamic empirical structural model of migration and educational attainment that retains the key components of our simplified model, but that tries to account for some of the salient features of the Chinese case as described in Section 3. In Section 6, the identification conditions and estimation procedure are discussed. Estimation results are presented in Section 7, and counterfactual simulations are presented in Section 8. Section 9 concludes.

2 Human Capital Accumulation and Migration: A Toy Model

In this section, we use a very simple model to illustrate that, provided a particular (but plausible) pattern of returns to education holds, the option value generated by the migration decision can lead to lower human capital accumulation for people at risk of migration. We only use this toy over-simplified model to illustrate our proposed mechanism. We defer the presentation and discussion of the full model to Section 5.

Consider first a world in which no migration is possible where an individual faces a simple binary decision of whether to invest in human capital or not. An individual in location r (rural), faces earnings of y_{rl} (l for low) if he does not invest in human capital, and of y_{rh} (h for high) if he

[2]Consistent with the finding in Zhou (2017) that social networks reduce migration costs.

does. Trivially, the individual will invest if $y_{rh} > y_{rl}$.

We now introduce a second location u (urban). For simplicity, we assume that an individual would choose to migrate from r to u if this were a possibility. We further assume that there is some exogenous probability P that an individual can migrate.[3] Let y_{ul} and y_{uh} denote earnings in the new location if an individual has low or high human capital, respectively. The individual's expected earnings when he does not invest in human capital are thus given by $Ey_l = y_{rl}(1-P) + y_{ul}P$, and the expected earnings when he gets an education are $Ey_h = y_{rh}(1-P) + y_{uh}P$. When migration is possible, an individual will invest in human capital provided that:

$$\begin{aligned} Ey_h > Ey_l &\iff y_{rh}(1-P) + y_{uh}P > y_{rl}(1-P) + y_{ul}P \\ &\iff \underbrace{[(y_{uh} - y_{rh}) - (y_{ul} - y_{rl})]}_{A} P + (y_{rh} - y_{rl}) > 0. \end{aligned} \tag{1}$$

Here the term in square brackets (A), measures the difference between the return to migrating when the individual invests in human capital and the return when he does not. When this term is negative, the probability of going to a higher level of education (i.e., $Pr(Ey_h - Ey_l > 0)$) will decrease. Alternatively, we can rewrite $A = (y_{uh} - y_{ul}) - (y_{rh} - y_{rl})$, and reinterpret our condition as stating that the return to human capital investment for migrants is smaller in urban than in rural areas.

The model we present here is too simplified to be of practical use. However, it is enough to illustrate the mechanism we want to highlight in this paper: if the return of moving to the city with a higher level of education is not as high as the return of moving to the city without it, individuals will be less likely to increase their human capital. This will be the case if, for example, a rural person moving to the city will get roughly the same job (say, as a cab driver or a cook) whether they are a high school graduate or not. In both cases the person will earn more in the city, but the *relative gain* is larger for the individual who did not invest in human capital.

The condition that the education return rate is lower in urban (for migrants) is different from what is usually assumed in most of the migration literature. Usually, researchers assume the return is higher (or the same) in the migration destination (e.g., Dustmann and Glitz, 2011). However, as we show below, this does not seem to be the case (at least at lower levels of education) for rural-urban migration in China.[4]

[3] The model of Section 5 allows for heterogenous endogenous migration probabilities.

[4] It is likely that a similar pattern holds for other cases, for example for Mexico-U.S. international migration (McKenzie and Rapoport, 2011).

3 Migration and Education in China: Background and Data

3.1 Education System in China

The education system in China consists of four different stages. The first stage is primary education (Grades 1 to 6).[5] The second stage is middle school (Grades 7 to 9), which completes the compulsory education level in China. As a consequence, both the central and the local government provide subsidies to schools to cover student tuition fees and other payments (e.g. textbooks, sports equipment).

The third stage is the high school period (senior secondary). There are two types of high schools: academic high schools and vocational high schools. Since students from both types of high schools are allowed to take the college entrance exam, we do not distinguish between these two types of high schools in our analysis. High school level education is not mandatory in China. Although there is an exam to sort students into different high schools, any individual can attend some type of high school, hence we do not consider whether there is a capacity constraint for high school admission in our analysis.

The fourth stage is college and beyond. After 1978, the government restarted the application of the college entrance exam to select students into college level education. Due to capacity constraints, the admission process is highly selective. Before 1980, less than 10% of individuals taking the exam gained admission into college. Before the year 1999, the total admission rate was still below 35%. Regarding tuition, until 1996 college tuition was very low as the government heavily subsidized college students. There are also "vocational" colleges (similar to community colleges in the US) where admission is not determined by the college entrance exam.

The big change in education policy came in 1999 when the college expansion policy started. Since 1999, there have been more than 60,000 additional admissions each year. Figure 1 shows that college admission offers have increased from 1.6 million to more than 6 million in ten years (1999-2009). As opposed to the migration channel described in Section 2, this expansion is likely to have increased the incentive to accumulate human capital by increasing the option value of schooling. In our analysis below, we also examine the impact of this policy on the education decisions of rural individuals.

3.2 Rural-Urban Migration in China

Since 1958, the Chinese central government has restricted the mobility of the population. From 1958 to 1978, mostly only the few rural people who had job offers in urban areas or recruitment letters from universities could migrate from rural to urban areas. In 1978 the people's commune system was replaced by the household-responsibility system, which led to a loosening of the re-

[5] This is true for most, but not all provinces. In some cases, for example Jiangxi before 2000, the primary school stage is five years instead of six.

strictions on rural residents mobility in 1983. After 1983 and until 1988, the central government lifted the restrictions on rural-urban migration. This migration policy was suspended between 1988 and 1990. After 1991, the government began to encourage rural-urban migration, and starting in 2000, the government started to reform the household registration system, or Hukou,[6] to further encourage more rural individuals to migrate. For example, in 2007, 12 provinces in China canceled the rural household registration, which meant that rural individuals had the same household registration as urban households in these provinces.[7] In these provinces, the local government does not distinguish between rural and urban residents any longer.

The easing of government restrictions on migration appears to have had a significant effect on people's migration decisions. There were 9.2 million people who migrated inter-province between 1990 and 1995 and this number increased to 32 million between 1995 and 2000, and to 38 million between 2000 and 2005. The stock of rural migrants increased from 78 million to 145 million within 10 years (see Figure 2).

3.3 Data

We use two main sources of data in our analysis: the first one is the China Household Income Project (CHIP) survey, a panel consisting of three waves from 2007 to 2009. The second source is the China Family Panel Study (CFPS). The CHIP dataset is designed to study issues such as the effect of rural-urban migration on income mobility and poverty alleviation, the state of education, and the health of children in migrating families. The survey consists of three representative samples of households, including a sample of rural households, a sample of rural migrant households, and a sample of urban households. For longitudinal 2007-2009 data, we mostly focus on the rural sample, which consists of individuals born in rural areas with a rural Hukou (household registration), and follow them over time, whether they migrate or not.[8]

We define a migrant as someone whose work is located in an urban area outside of the county of their rural Hukou registration location. Tables 1 and 2 present some basic summary statistics for the data we employ.[9] The sample contains information on work experience, work location, earnings, and education choices. Using this data, we can construct the migration histories, education decisions, location choices, and work status for individuals who are between 16 and 60 years old. We restrict our attention to males with a rural household registration in order to simplify

[6]A household registration record officially identifies a person as a resident of an area and includes identifying information such as name, parents, spouse, and date of birth.

[7]These 12 provinces are Chongqing, Fujian, Guangxi, Hebei, Hubei, Hunan, Jiangsu, Liaoning, Shandong, Shanxi, Sichuan and Zhejiang.

[8]There are two main reasons why we do not use the migration sample. First, the response rate in the migration data is quite low. The attrition rate is above 70% for the three years of panel data. Second, we cannot follow the history of migrants' work experience using the migration samples. For example, migrants who return to their hometowns are not surveyed in the migration sample. In comparison, the attrition rate for the rural sample is less than 1% over the three years.

[9]We deflate earnings to 1995 Yuan using regional price indices, which we construct following Brandt and Holz (2006).

the structural model we develop in Section 5. As we show in Appendix Tables A1-A3, the same patterns we illustrate in this section hold even more strongly for females.

Table 1 compares demographics between individuals with urban Hukou and those with rural Hukou. It is clear that the urban group has higher years of education than that of the rural group (i.e., 12.34 years vs. 9.22 years). They are also three times as likely to take the college entrance exam than the rural group. However, there are no significant differences for non-cognitive skill measures between these two groups.

Table 2 is restricted to individuals with a rural Hukou and compares migrants versus non-migrants. On average, both earnings and years of education are higher for rural migrants than for non-migrants. Self-reported class performance is slightly better for non-migrants. The fraction taking the collage entrance exam, on the other hand, is very close between migrants and non-migrants.

Table 3 provides summary statistics for the rural sample we use for constructing auxiliary regressions and moments in the structural model. In the sample, the average years of education is about 9 years. More than 70% of the sample just graduates from middle school. About 10% of rural individuals go on to some type of college level education. The mean of real log monthly earnings is quite stable over the three years, as is the fraction migrating.

The average year of the first time migration in 2007 is 1999. When focusing on this measure across different cohorts, we find that for the cohort born before 1960, the mean first time migration year is about 1993 with a large standard deviation. For the people born after 1980, the mean of first time migration is 2002, and most of them migrated just after mandatory education. When looking at the fraction of rural individuals who ever migrate across different cohorts, we find that younger cohorts tend to migrate more. For example, only 25% of older cohorts migrate but around 84% of younger cohorts have experienced migration.

In order to have better measures for individual cognitive and non-cognitive skills, we supplement the CHIP data with the first three (biannual) waves (2010-2014) of the China Family Panel Study (CFPS). We use both CFPS and CHIP data to estimate the distributions of individual unobserved skill endowments as in Heckman and Navarro (2007) and Navarro and Zhou (2017) as described in Section 6. We group data measures into three different categories: cognitive skill measures, non-cognitive skill measures, and measures for both skills. Some of the measures we observe repeatedly over time.

As a measurement system for cognitive skills, we include the word, math, and word recall tests from the CFPS data, and from CHIP we include whether the individual takes the college entrance exam, as well as a self- reported measure of class performance. For non-cognitive skills, from CHIP we use five measures related to how easy it is for the individual to concentrate to do things, how capable is the individual of making decisions, how capable to face problems and overcome difficulties, and whether they lack self-confidence. Our non-cognitive measures from CFPS data include measures of how easy it is for the individual to concentrate to do things, how well the individual gets along with others, whether the individual often feels upset, and whether the individual finds things difficult all the time. Finally, we assume that smoking (included in both datasets) and drinking (only CFPS) behaviors are functions of both cognitive and non-cognitive skills.

Table 4 gives summary statistics for the measures in the sample of CHIP data we use for estimation of the skills model. In this sample, the average years of education is close to 8 years. The self reported class performance is concentrated at the normal category. Only 7% of rural individuals have taken the college entrance exam. For non-cognitive skill measures, most people have positive attitudes about their ability to concentrate to do things, making decisions, and overcoming difficulties. There are 70% of people who ever smoke, and they smoke around 10 cigarettes per day.

Table 5 shows summary statistics for the CFPS data. We use six cognitive skill measures: two math, two language, and two word recall test measures. Regarding non-cognitive skills, we have four measures. As with the CHIP data, rural individuals have very positive evaluations about their skills. In terms of smoking and drinking behaviors, the pattern is similar to CHIP: a large fraction of rural males smoke in China. More than 50% of rural males drink more than 500g alcohol per week, and 25% of them drink more than 1000g per week.

4 Education and Rural-Urban Migration in China: Preliminary Evidence

In this section, we illustrate that the pattern described in the previous section, both in terms of returns to education and migration, seems to be present in China. We run Mincer-type regressions in Table 6. As expected, there seems to be a positive return to migration, even after controlling for education. Furthermore, the return to education in rural areas is between 5% and 6%, consistent with the findings in Heckman and Li (2004). For both years, we find that the difference in the return to education for rural individuals in urban areas compared to rural areas (the A term in equation 1) is negative, consistent with the pattern described in our toy model in the previous section. As we show in columns 2 to 4 of the table, the pattern holds even after we control for different measures of cognitive and non-cognitive skills.

In order to investigate whether this pattern has led to a decline in educational attainment associated with migration, we take advantage of a change in migration policy that happened in the early 1980's in China. Migration in China has been restricted by the government since 1958. In 1978, the "people's commune system" was replaced by the "household-responsibility system", which loosened the restrictions on rural residents mobility. However, the central government still kept strict restrictions limiting the opportunities of working in cities until around 1982-1983. As a consequence, we consider the policy as having taking place in 1983.

We begin by graphically analyzing in Figure 3 whether the elimination of restrictions to migration has an effect consistent with the predictions of our model in Section 2. In the figure we plot the evolution of average years of education for both individuals with an urban and a rural Hukou. In order to highlight the patterns we wish to illustrate, we add 1.25 to the rural average educational attainment to make sure that the lines coincide pre-1982. As the figure shows, the educational attainment of both groups follows closely until around 1982, at which point rural educational attainment slows down considerably relative to its urban counterpart. This pattern continues to hold over the next decades, and it is not until around 2003, four years after the massive college expansion program started in 1999, that the pattern seems to break.

We more formally investigate whether the pattern we show in Figure 3 holds by running a difference-in-differences specification for educational attainment before and after 1983. We begin by investigating whether the parallel trends assumption pre-1983 holds as a way of providing some supporting evidence for post-1983 parallel trends holding as needed for a DiD comparison to be valid. In Table 7 we run a regression of years of education against an indicator for rural, year dummies up to 1982, and interactions. As we can see, the interactions (i.e., the difference in the trend for urban vs rural) is not significant, so we cannot reject the hypothesis of parallel trends. This pattern, that the interaction of the year dummy and rural is not significant, holds even after we control for measures of cognitive and non-cognitive skills.

We next run a DiD specification to examine the effect that the loosening of the migration restrictions had on years of education in Table 8. Let $e_{i,t}$ denote years of education for individual i at time t, $Rural_i$ be an indicator for i having a rural Hukou, and $\mathbb{1}_{>1982_t}$ be an indicator for $t > 1982$. The specification we run is given by

$$e_{i,t} = \beta_0 + Rural_i\beta_1 + \mathbb{1}_{>1982_t}\beta_2 + \mathbb{1}_{>1982_t} \times Rural_i\beta_3 + X_{i,t}\beta_4 + \varepsilon_{i,t}$$

where $X_{i,t}$ denotes additional controls as specified in Table 8 for different columns. The coefficient on the interaction term, β_3, reflects the impact of the changes in migration policy on the individual's schooling choices.

As shown in the table, after 1982 rural individuals have significantly less education, between 0.4 and 0.3 years less, than the pre-1983 pattern implies, consistent with the predictions of our model. This is true even after we control for different measures of cognitive and non-cognitive skills. We restrict our specification to include years up to 1989 in order to isolate the effect of migration policy from other policies that happened after. As we show in Table A4 in the Appendix, the negative effect is even larger if we include later years.

In order to further isolate the effect of the changes in migration policy from other changes that may have happened during the same years, we repeat the DiD exercise where we attempt to isolate individuals who were more or less likely to have been affected by the policy (i.e., a triple difference). We first run a probit model for migration including only pre-1983 data, i.e., data before the policy was implemented. We then use the estimates from this pre-1983 model to predict the probability of migration for post-1983 data. The idea behind this exercise is to identify the individuals who were are the margin for migrating even if the policy had not been implemented. We then use this predicted probability to classify individual into more or less likely to migrate. If the migration policy is the main force driving our results, we would expect the decline in human capital investment to be even larger for marginal individuals. The results for different classifications are presented in Table 9. As expected, depending on the classification, the policy is predicted to have reduced average years of education in rural China between 0.13 and 0.9 years.

5 Structural Model

The evidence in Section 4 shows that, consistent with the mechanism we suggest in Section 2, the existence of a rural-urban migration option can reduce the incentive to invest in human capital for rural individuals. The evidence, however, rests on us arguing that migration became easier after 1982. Furthermore, several other changes, like the massive expansion of the college system in 1999 were put in place during the period we study. To further analyze whether our hypothesized mechanism is responsible for the pattern we observe, to quantify the magnitude of the migration option on rural education choices, and to parse the effects of the college expansion policy on rural individuals decisions, we now develop a formal structural model of education and migration decisions in China which we take to the data in Section 6.

Let $x_{i,a}$ be the state vector for an individual i, who is of age a. We assume that $x_{i,a}$ contains, among other things, a vector of individual endowments $\theta_i = \left(\theta_i^C, \theta_i^N\right)'$ (cognitive and non-cognitive skills) that is unobserved to the econometrician. It also contains the calendar year the individual is at, t. Let $d_{i,a} = k$ if an individual makes choice k, where $k = s$ if the individual chooses to attend school, $k = r$ if he instead works in a rural area (i.e., stays in rural or return migrates to rural), and $k = u$ if he chooses to work in an urban area (i.e., stay in urban or migrate to urban). We assume that once an individual drops out of school he does not return which is consistent with what we observe in the data. Consequently, the choice set consists of $\{s, r, u\}$ if $d_{i,a-1} = s$, and of $\{r, u\}$ if $d_{i,a-1} \neq s$. Figure 4 provides a representation of the timing/nodes at which different choices are available.

The utility flow for individual i associated with choice $d_{i,a} = k$ is specified as $u(x_{i,a}, k) + \xi_{i,a}^k$, where $\xi_{i,a}^k$ is a random variable unobserved to the econometrician. For computational simplicity, we assume that $\xi_{i,a}^k$ is distributed extreme value type-I, i.i.d. across locations and across periods, and independent of $x_{i,a}$. We let $\xi_{i,a} = \left(\xi_{i,a}^s, \xi_{i,a}^r, \xi_{i,a}^u\right)'$ if $d_{i,a-1} = s$, and $\xi_{i,a} = \left(\xi_{i,a}^r, \xi_{i,a}^u\right)'$ if $d_{i,a-1} \neq s$.

5.1 Flow Payoffs: Urban

Let $e_{i,a}$ denotes the years of education an individual has acquired up to a, and $\exp^r_{i,a}, \exp^u_{i,a}$ denote accumulated work experience in rural and urban areas respectively. Individuals working in a city receive earnings according to

$$\ln y^u_{i,a} = \gamma^u_0 + e_{i,a}\gamma^u_1 + \exp^r_{i,a}\gamma^u_2 + \left(\exp^r_{i,a}\right)^2\gamma^u_3 + \exp^u_{i,a}\gamma^u_4 + \left(\exp^u_{i,a}\right)^2\gamma^u_5 + \mathbb{1}_{i,a,col}\gamma^u_6 + \theta'_i\gamma^u_7 + \varepsilon^u_{i,a}, \quad (2)$$

where $\varepsilon^u_{i,a} \sim N\left(0, \sigma^2_{\varepsilon^u}\right)$ is an i.i.d. shock to earnings. We also allow for an extra college premium in the earning equation. $\mathbb{1}_{i,a,col}$ is an indicator that takes value one if individual i at age a is a college graduate. In China, individuals who want to go to college need to take the college entrance exam. Before 1999 less than 35% of individuals who took the exam got a college offer. Therefore, we expect there exists an extra college wage premium γ^u_6.

An individual moving from a rural area last period to an urban area this period pays a migration cost given by

$$mc_{i,a,t} = \gamma^m_0 + t\gamma^m_1 + a\gamma^m_2 + a^2\gamma^m_3 + \theta'_i\gamma^m_4 + \mathbb{1}_{t>1982}\gamma^m_5 + \mathbb{1}_{t>1982}t\gamma^m_6 + \mathbb{1}_{t>1989}\gamma^m_7 + \mathbb{1}_{t>1989}t\gamma^m_8, \quad (3)$$

where we allow for a trend in migration costs (γ^m_1). If $\gamma^m_1 < 0$, it would reflect the fact that migration has become easier over time in China (e.g., the improvement of transportation conditions). We further include an indicator for the post-1982 period (γ^m_5, and γ^m_6), to account for the policy change described in Section 3. During the year 1988-89, the central government temporary prohibited migration. After 1990, the government began to promote migration again. To capture these policy changes, we want to separately estimate γ^m_7, and γ^m_8. The flow utility of an individual in an urban area is thus given by

$$u\left(x_{i,a}, u\right) + \xi^u_{i,a} = \ln y^u_{i,a} - mc_{i,a,t}\mathbb{1}_{d_{i,a-1}\neq u} + \xi^u_{i,a}. \quad (4)$$

5.2 Flow Payoffs: Rural

Individuals working in a rural area receive earnings according to

$$\ln y^r_{i,a} = \gamma^r_0 + e_{i,a}\gamma^r_1 + \exp^r_{i,a}\gamma^r_2 + \left(\exp^r_{i,a}\right)^2\gamma^r_3 + \exp^u_{i,a}\gamma^r_4 + \left(\exp^u_{i,a}\right)^2\gamma^r_5 + \mathbb{1}_{col}\gamma^r_6 + \theta'_i\gamma^r_7 + \varepsilon^r_{i,a}, \quad (5)$$

with $\varepsilon^r_{i,a} \sim N\left(0, \sigma^2_{\varepsilon^r}\right)$. Besides deriving utility from earnings, individuals in rural areas also derive extra utility from being in a rural location (a "home premium", e.g., Kennan and Walker, 2011) given by

$$h_{i,a} = \gamma^h_0 + a\gamma^h_1 + a^2\gamma^h_2 + \theta'_i\gamma^h_3. \quad (6)$$

Finally, an individual moving back to a rural area, pays a return migration cost. For identification purposes, we restrict the return migration cost function to have the same coefficients as (3) for the variables which are not affected by the migration policies

$$\varphi_{i,a,t} = \gamma_0^m + t\gamma_1^m + a\gamma_2^m + a^2\gamma_3^m + \theta_i'\gamma_4^m. \tag{7}$$

The flow utility of an individual in a rural area is thus given by

$$u\left(x_{i,a}, r\right) + \xi_{i,a}^r = \ln y_{i,a}^r + h_{i,a} - \varphi_{i,a,t}\mathbb{1}_{d_{i,a-1}=u} + \xi_{i,a}^r. \tag{8}$$

5.3 Flow Payoffs: Schooling

An individual derives utility from attending school that depends, in part, on the schooling level that he is attending:

$$\begin{aligned} u\left(x_{i,a}, s\right) + \xi_{i,a}^s = & \left(\gamma_{1,hs}^s + t\gamma_{2,hs}^s + \mathbb{1}_{1965<t<1977}\gamma_{3,hs}^s + \theta_i'\gamma_{4,hs}^s\right)\mathbb{1}_{hs} \\ & + \left(\gamma_{1,voc}^s + t\gamma_{2,voc}^s + \mathbb{1}_{1965<t<1977}\gamma_{3,voc}^s + \theta_i'\gamma_{4,voc}^s\right)\mathbb{1}_{voc} \\ & + \left(\gamma_{1,col}^s + t\gamma_{2,col}^s + \mathbb{1}_{1965<t<1977}\gamma_{3,col}^s + \theta_i'\gamma_{4,col}^s\right)\mathbb{1}_{col} + \xi_{i,a}^s, \end{aligned} \tag{9}$$

where $\mathbb{1}_{hs}$ is an indicator that takes value 1 if the individual is attending high school, and similarly for *voc* which stands for vocational college (similar to the US community college) and *col* for college. Since it is very uncommon for students to work in China, we do not allow for part-time work while in school. $\mathbb{1}_{1965<t<1977}$ indicates whether the education choices were taken during the Cultural Revolution period. During that period, the government expanded secondary education rapidly but destroyed both the college and vocational education system. In the model, we allow individuals to evaluate schools differently between Cultural Revolution periods and non-Cultural Revolution periods.

5.4 Individual Choices

The value that an individual gets from making a particular choice can be defined recursively as follows. Let

$$V_a^k\left(x_{i,a}, \xi_{i,a}^k\right) = u(x_{i,a}, k) + \xi_{i,a}^k + \beta E\left[V_{a+1}(x_{i,a+1}, \xi_{i,a+1}) | x_{i,a}, d_{i,a} = k\right]$$

be the value that an individual gets if he chooses to attend school ($k = s$), or work on either $k = r$ or $k = u$, and then continues to maximize utility every period, discounted using β.

First, consider an individual who is done with schooling. He will choose such that

$$d_{i,a} = \arg\max_{k\in\{r,u\}}\left\{V_a^r\left(x_{i,a}, \xi_{i,a}^r\right), V_a^u\left(x_{i,a}, \xi_{i,a}^u\right)\right\},$$

and

$$V_a\left(x_{i,a},\xi_{i,a}\right)=\max\left\{V_a^r\left(x_{i,a},\xi_{i,a}^r\right),V_a^u\left(x_{i,a},\xi_{i,a}^u\right)\right\}.$$

Since this is a lifecycle model, individuals will face a final period A, which we set to age 60. Since we do not model what happens after retirement, etc, we simply model this terminal value as a function of the states at that point. In particular, we normalize

$$V_A\left(x_{i,A},r\right)=0+\xi_{i,A}^r,$$

and write

$$V_A\left(x_{i,A},u\right)=\alpha_0+e_{i,a}\alpha_1+\exp_{i,a}^r\alpha_2+\left(\exp_{i,a}^r\right)^2\alpha_3+\exp_{i,a}^u\alpha_4+\left(\exp_{i,a}^u\right)^2\alpha_5+\theta_i'\alpha_6+t\alpha_7+\xi_{i,A}^u.$$

Next, consider an individual who is already enrolled in either College or Vocational school. While we assume that the $x_{i,a}$ contains an indicator of the schooling level that the individual is attending, we abuse notation and write $x_{ia}=(\tilde{x}_{i,a},\ell)$ for $\ell=\{hs,voc,col\}$, to make it explicit. We thus say that the value of going to school in college is given by $V_a^s\left(\tilde{x}_{i,a},col,\xi_{i,a}^s\right)$ to recognize that his flow utility will be based on $\gamma_{1,col}^s+\theta_i'\gamma_{2,col}^s+\mathbb{1}_{1965<t<1977}\gamma_{3,col}^s+\theta_i'\gamma_{4,col}^s+$ in equation 9. Similarly for a vocational student we use $V_a^s\left(\tilde{x}_{i,a},voc,\xi_{i,a}^s\right)$ to denote the value of going to school in this case. For an individual in college, his decision every period will be given by

$$d_{i,a}=\arg\max_{k\in\{s,r,u\}}\left\{V_a^s\left(\tilde{x}_{i,a},col,\xi_{i,a}^s\right),V_a^r\left(x_{i,a},\xi_{i,a}^r\right),V_a^u\left(x_{i,a},\xi_{i,a}^u\right)\right\}, \tag{10}$$

his value by

$$V_a\left(x_{i,a},\xi_{i,a}\right)=\max\left\{V_a^s\left(\tilde{x}_{i,a},col,\xi_{i,a}^s\right),V_a^r\left(x_{i,a},\xi_{i,a}^r\right),V_a^u\left(x_{i,a},\xi_{i,a}^u\right)\right\}, \tag{11}$$

and similarly for vocational school.

Now, consider the decision to attend college. An individual gets to decide whether to attend college only if he is admitted into one. We assume that the probability of getting a college offer is given by:

$$\lambda_{i,t}^{col}=\frac{\exp(\gamma_0^{col}+\theta_i'\gamma_1^{col}+t\gamma_2^{col}+t\mathbb{1}_{t\geq1999}\gamma_3^{col}+\mathbb{1}_{t\geq1999}\gamma_4^{col})}{1+\exp(\gamma_0^{col}+\theta_i'\gamma_1^{col}+t\gamma_2^{col}+t\mathbb{1}_{t\geq1999}\gamma_3^{col}+\mathbb{1}_{t\geq1999}\gamma_4^{col})}, \tag{12}$$

where γ_3^{col}, and γ_4^{col} are included to account for the massive sustained increase in the number of college offers that happened since 1999, as shown in Figure 1. Hence, the value for an individual who has a college offer, will be

$$V_a\left(x_{i,a},\xi_{i,a}\right)=\max\left\{V_a^s\left(\tilde{x}_{i,a},col,\xi_{i,a}^s\right),V_a^s\left(\tilde{x}_{i,a},voc,\xi_{i,a}^s\right),V_a^r\left(x_{i,a},\xi_{i,a}^r\right),V_a^u\left(x_{i,a},\xi_{i,a}^u\right)\right\}. \tag{13}$$

Individuals who do not get a college offer, may still attend a vocational college. An individual who does not have a college offer (either because he did not get one, or because he did not take the college entrance exam) will decide based on

$$V_a\left(x_{i,a},\xi_{i,a}\right)=\max\left\{V_a^s\left(\tilde{x}_{i,a},voc,\xi_{i,a}^s\right),V_a^r\left(x_{i,a},\xi_{i,a}^r\right),V_a^u\left(x_{i,a},\xi_{i,a}^u\right)\right\}. \tag{14}$$

Next, consider someone who is deciding whether to enroll in their last year of high school. The timing is as follows. An individual first decides whether to enroll in the last year of high school before he gets to observe the cost of the entrance exam. Later in the same year the cost of the exam is realized and the individual decides whether to take the exam. If he takes the college exam, then at the beginning of the next year he draws according to Equation (12) and decides based on Equation (13). If he does not have a college offer then he chooses based on Equation (14).

If the individual decides to take the college entrance exam, he has to pay the (psychic) cost of the college entrance exam, given by

$$ce_i=\gamma_0^{ce}+\theta_i^{'}\gamma_1^{ce}+\varepsilon_i^{ce}.$$

The decision of whether to take the college entrance exam is based on his expected value of attending either college or vocational college, where the expectation is taken with respect to $\xi_{i,a},\varepsilon_{i,a}^r,\varepsilon_{i,a}^u$. If we let

$$\begin{aligned} E\mathcal{V}^{col}\left(x_{i,a}\right) &= E_a\max\{V_{a+1}^s\left(\tilde{x}_{i,a+1},col,\xi_{i,a+1}^s\right),V_{a+1}^s\left(\tilde{x}_{i,a+1},voc,\xi_{i,a+1}^s\right), \\ &\quad V_{a+1}^r\left(x_{i,a+1},\xi_{i,a+1}^r\right),V_{a+1}^u\left(x_{i,a+1},\xi_{i,a+1}^u\right)\}, \end{aligned} \tag{15}$$

and

$$\begin{aligned} E\mathcal{V}^{voc}\left(x_{i,a}\right) &= E_a\max\{V_{a+1}^s\left(\tilde{x}_{i,a+1},voc,\xi_{i,a+1}^s\right),V_{a+1}^r\left(x_{i,a+1},\xi_{i,a+1}^r\right), \\ &\quad V_{a+1}^u\left(x_{i,a+1},\xi_{i,a+1}^u\right)\}, \end{aligned} \tag{16}$$

then, he will take the exam if

$$\beta\lambda_{i,t}^{col}E\mathcal{V}^{col}\left(x_{i,a}\right)+\beta\left(1-\lambda_{i,t}^{col}\right)E\mathcal{V}^{voc}\left(x_{i,a}\right)-ce_i>\beta E\mathcal{V}^{voc}\left(x_{i,a}\right).$$

Let

$$\mathcal{V}^{hs}\left(x_{i,a}\right)=E\max\left\{\beta\lambda_{i,t}^{col}E\mathcal{V}^{col}\left(x_{i,a}\right)+\beta\left(1-\lambda_{i,t}^{col}\right)E\mathcal{V}^{voc}\left(x_{i,a}\right)-ce_i,\beta E\mathcal{V}^{voc}\left(x_{i,a}\right)\right\},$$

that is the value of attending high school before the individual gets to observe the cost of taking the exam, i.e., before ε_i^{ce} is realized. The value of enrolling in the last year of high school for this individual is

$$V_a^s\left(\tilde{x}_{i,a},hs,\xi_{i,a}^s\right)=u(\tilde{x}_{i,a},hs,s)+\xi_{i,a}^s+\mathcal{V}^{hs}\left(x_{i,a}\right).$$

Hence, the individual will enroll in the last year of high school if

$$V_a^s\left(\tilde{x}_{i,a}, hs, \xi_{i,a}^s\right) > \max\left\{V_a^r\left(x_{i,a}, \xi_{i,a}^r\right), V_a^u\left(x_{i,a}, \xi_{i,a}^u\right)\right\},$$

and his value will be

$$V_a\left(x_{i,a}, \xi_{i,a}\right) = \max\left\{V_a^s\left(\tilde{x}_{i,a}, hs, \xi_{i,a}^s\right), V_a^r\left(x_{i,a}, \xi_{i,a}^r\right), V_a^u\left(x_{i,a}, \xi_{i,a}^u\right)\right\}.$$

Finally, in any previous year, the choice for an individual deciding whether to attend (or remain) in high school is

$$d_{i,a} = \arg\max_{k\in\{s,r,u\}}\left\{V_a^s\left(\tilde{x}_{i,a}, hs, \xi_{i,a}^s\right), V_a^r\left(x_{i,a}, \xi_{i,a}^r\right), V_a^u\left(x_{i,a}, \xi_{i,a}^u\right)\right\},$$

and his value is given by

$$V_a\left(x_{i,a}, \xi_{i,a}\right) = \max\left\{V_a^s\left(\tilde{x}_{i,a}, hs, \xi_{i,a}^s\right), V_a^r\left(x_{i,a}, \xi_{i,a}^r\right), V_a^u\left(x_{i,a}, \xi_{i,a}^u\right)\right\}.$$

6 Identification and Estimation

6.1 Estimation

We use indirect inference to estimate our model. There are two sets of parameters we need to estimate. One set of parameters is the density of unobserved skill endowments, Θ_1 as described in the next section. The second set of parameters are those remaining in the structural model Θ_2.

The estimation procedure is as follows:

1. First, we estimate Θ_1 using the two types of skill measures in the data via indirect inference.
2. Then, given Θ_1, for each guess of Θ_2, we solve the dynamic discrete choice model recursively.
3. Given the solution of the model from Step 2, we simulate individuals' decisions and get auxiliary parameters and moments to form the indirect inference objective function, and iterate the optimization until convergence.
4. Since estimation of Θ_2 was done given estimates of Θ_1 obtained in a first stage, we take one last Newton step for both Θ_1 and Θ_2 together to correct standard errors.
5. We perform the procedure on various simulated datasets to make sure the auxiliary models chosen for indirect inference contain the necessary information to identify the parameters of the model.

6.1.1 Skill Distribution Estimation

For the case in which we have continuous measures, $M_{i,t,j}$ $j = 1, ..., J$, we model the unobserved skills as being related to the measures using a linear model where we control for age and years of education at test date:

$$M_{i,t,j} = X'_{i,t}\lambda^X_{t,j} + \theta'_i\lambda^\theta_{t,j} + u_{i,t,j}.$$

Depending on whether the measure is considered to be associated with cognitive, non-cognitive or both skills, some component of $\lambda^\theta_{t,j}$ may be normalized to zero. We use a similar equation as auxiliary model for the indirect inference procedure

$$M_{i,t,j} = X'_{i,t}\pi^X_{t,j} + \omega_{i,t,j}$$

as well as moments on the joint distribution of the estimated $\omega_{i,t,j}$.

For the case of discrete skill measures, we relate the skills to the measures using an ordered model of the form:

$$M_{i,t,j} = \begin{cases} 1, & \text{if} \quad X'_{i,t}\lambda_{X,t,j} + \theta'_i\lambda_{\theta,j} + u_{i,t,j} < m_{1,t,j} \\ k, & \text{if} \quad m_{k-1,t,k} \leq X'_{i,t}\lambda_{X,t,j} + \theta'_i\lambda_{\theta,t,j} + u_{i,t,j} < m_{k,t,j} \text{ for } 1 < k < K \\ K, & \text{if} \quad X'_{i,t}\lambda_{X,t,j} + \theta'_i\lambda_{\theta,t,j} + u_{i,t,j} > m_{K-1,t,j} \end{cases}. \tag{17}$$

We use the following OLS regression as the auxiliary regression:

$$\mathbb{1}_{M_{i,t,j}=1} = \sum_{k=2}^{K} \pi^M_{k,t,j}\mathbb{1}_{M_{i,t,j}=k} + X'_{i,t}\pi^X_{t,j} + \omega_{i,t,j}, \tag{18}$$

as well as moments on the joint distribution of the estimated $\omega_{i,t,j}$.

We parametrize the distribution of θ to follow a mixture of 2 independent normals. Namely, we let

$$\theta^{cog}_i \sim MN\left(\mu^{cog}_1, \mu^{cog}_2, \sigma^{cog}_1, \sigma^{cog}_2, p\right)$$

where we normalize the overall mean $p\mu^{cog}_1 + (1-p)\,\mu^{cog}_2 = 0$. Similarly for non cognitive skills we let

$$\theta^{ncog}_i \sim MN\left(\mu^{ncog}_1, \mu^{ncog}_2, \sigma^{ncog}_1, \sigma^{ncog}_2, p\right)$$

and normalize the mean to zero as well. Notice, however, that even though the component normals are independent, we correlate the two skills by assuming that for any i both come from either the first set of normals (with probability p) or the second one. We parametrize all the uniquenesses (the $u's$) to follow normals with mean zero and variance to be estimated (or normalized to one for the ordered measures). Finally, for identification purposes, we normalizes the loading (i.e., the λ) for cognitive skills to one in a math test measure from CFPS data, and similarly for the non-cognitive we normalize the loading on θ^{ncog} to one for the measure of "getting on well with others".

6.1.2 Structural Model Estimation

For the structural model, we choose three wage regressions for the indirect inference auxiliary model: the first wage regression is comparable with the reduced form analysis. We pool migrants and non-migrants together and control for age, years of education, migration status, the interaction term of migration times years of education, college graduation indicator, and cognitive and non-cognitive skills, θ. The other two wage regressions are the wage regression for those who never migrate, and for the individuals for whom we can observe the full migration history.

We use two migration and two return migration regressions to pin down migration and return migration decisions. There are four regressions for different education level choices, where we use binary indicators for each education level choice. In all cases we include moments on the distribution of the residuals as well as the regression coefficient estimates. We also include other additional moments for different cohorts: the first year of migration and the fraction who ever migrated by 2007 across four cohorts.

6.2 Identification

In this section, we briefly sketch how identification of all the parameters in the structural model is obtained.

First, to identify the joint distribution of unobserved skills, we use two different data sets which include both cognitive and non-cognitive measures of individual's skills. Following the method of Cunha, Heckman, and Schennach (2010b), we can identify the joint distribution of the two different unobserved skills.

Second, we assume that the location specific log earning distributions are normals. Therefore, the observed log earnings identify the location specific earning function and the variances of the earning shocks $\sigma^2_{\epsilon^u}$, and $\sigma^2_{\epsilon^r}$.

Third, to identify the location amenity value, migration and return migration costs, since we only can identify the relative location amenity value, we normalize the location amenity value of living in urban areas equals to zero. Then we follow the Kennan and Walker (2011) identification strategy to separately identify the location amenity value and migration costs.

Fourth, since we observe individuals drop out from different levels of schooling, these choices help us to identify the level of flow utility at different schooling levels. Given the flow utility level at different schooling levels, the probability of getting college offer, $\lambda^{col}_{i,t}$ can be identified by comparing the fraction of individuals who gets college offers and the fraction of individuals who taking the college entrance exam.

Identification of the other components of our model, follows directly from the analysis of Heckman and Navarro (2007), and French and Taber (2011).

7 Estimation Results

7.1 Model Fit

As we described in Section 6, we use indirect inference to estimate the skill distributions first, and then estimate the structural model. Tables 10-12 and A5 to A11 in the Appendix present the results of the auxiliary models we used as well as the corresponding equivalent parameters generated from our estimated model. If we perform a joint test of equality between data and model estimates, we get a $\chi^2_{368} = 2.47$ which shows that the model estimates fit the data really well. As an example, Table 10 shows the auxiliary regressions for the log earnings model. Almost all model estimates are within one standard deviation of the estimates from the original data.

Tables 10-12 give the auxiliary model regression parameters and moments for the structural model. The model joint moments goodness fit test gives a $\chi^2_{130} = 39.55$ which produces a p-value close to 1. We cannot reject the equality of model and data estimates for these regressions either.

7.2 Estimation Results

In Tables 13-15 (and A12 in the Appendix) we present the estimated parameters for the model. In Table 13, we show the estimates for rural and urban earning equations. First, thing to notice is that, as expected, the baseline average earnings of a person with rural Hukou are much higher in urban areas (i.e., a constant term of 6.05 vs 5.53). Second, the education return rate is higher in rural areas than in urban areas for individuals with rural Hukou. In rural areas, the education rate of return is around 0.036, and the return in urban areas is around 0.015 for rural migrants. The rural education return rate is more than double that in urban areas. The college graduate premium, however, is higher in urban areas for rural migrants. These findings regarding the differences for the education returns and the constant term in rural and urban areas, support our regression analysis of Section 4, in particular that the term A in Equation (1) is negative.

Cognitive and non-cognitive skills for individuals with a rural Hukou are rewarded differently in rural and urban areas. In rural areas, the returns to both cognitive and non-cognitive skills are negligible. In urban areas on the other hand, rural migrants face a large and positive return to cognitive skills but a negative return to non-cognitive skills.

Table 14 gives the estimates for migration costs. As expected, the changes in migration policy implemented in 1982-83 are reflected in a significant reduction of migration costs in the model. Furthermore, the model also captures the decreasing migration cost trend over time consistent with the further pushes to increase migration by the Chinese government as well improvements in the transportation system. We also find that both cognitive and non-cognitive skills, but especially the later, have a negative impact on migration costs, reflecting that social skills can help rural individuals when migrating to urban areas, consistent with the findings in Zhou (2017).

In Table 15, we present estimates of the utility parameters for different levels of education. In general, individuals with high cognitive skills have higher utility values for schooling. Interestingly, non-cognitive skills have a large negative impact on the psychic cost of taking the college entrance exam.

7.3 Selection

In order to understand the patterns of selection that migration generates, in Figure 5 we present distributions of earnings, both factual and counterfactual ones. Panel A in the figure presents the distribution of earnings conditional on migration. Since a migrant moves to the city, the distribution of urban earnings conditional on migration (the dashed line) is a factual one, while the sold line presents the counterfactual distribution of earnings that migrants would have obtained had they stayed in the rural area. What the figure clearly shows is that migrants are positively selected in terms of earnings, i.e., the people who migrate are those that get higher earnings than what they could get if they stayed. Similarly, in Panel B we present the earnings distributions for non-migrants, where the pattern also shows that the people who do not migrate get higher earnings in the rural area (dashed line) than what they could have gotten had they chosen to migrate (the solid line).

To further investigate how selection is affected by the cognitive and non-cognitive skills that the individuals have, and how this patterns has changed, in Figures 6 and 7 we present the distributions of skills for different cohorts according to when their first migration episode happened. Figure 6 presents the results for cognitive skills. The first thing to notice is that people who were able to migrate before the policy reform have high cognitive skills. This is not the case for people who migrate early in their life post-policy, as they tend to be low cognitive skills individuals. The final thing to notice is that, while the pattern is bimodal, people who never migrate tend to be high cognitive skills individuals as well.

Figure 7 repeats the exercise for non-cognitive skills where similar patterns emerge. As with cognitive skills, people who managed to migrate before the policy have high non-cognitive skills. Similarly, while the pattern is bimodal, people who never migrate tend to have high non-cognitive skills as well.

8 Counterfactual Policy Simulations

Table 16 presents the results of performing counterfactual policy simulations on individuals' education and migration choices. We first analyze the effect of eliminating the migration policy of 1982 by setting the parameters associated with the change in migration costs post-1982 to zero. The results are presented in the column labeled "No Migration Policy". The first thing to notice is that, as expected, if migration restrictions had not been relaxed the average level of education would have increased by 0.23 years, which is about 72% of the 0.32 years predicted by our DiD exercise. Most of this change comes from more individuals attending high school and vocational college. So, while our migration mechanism does not capture the whole effect, it accounts for almost 3/4 of the estimated change.

In Table 17 we decompose the transitions across education levels that happen as a result of our no migration policy. As the table shows, most of the action comes from having people who originally would stop at middle school (around 70% of the sample) now deciding to attend high school (6.35%) and a few attending vocational school. Similarly, a large proportion of people who originally attend high school (about 21% of the sample) now choose to attend vocational school (23.22%).

Having established that the elimination of the restrictions to migration has resulted in lower educational attainment by rural individuals, we next turn our attention to policies that may ameliorate this effect. In particular, we investigate whether the unprecedented college expansion that the Chinese government began in 1999 has been enough to compensate for the decline in human capital accumulation that we observe. In the column labeled "No College Expansion" in Table 16, we set the relevant parameters in the admission offer probability equal to zero to eliminate the impact of the expansion policy. When we eliminate the college expansion policy, average years of education declines from 9.11 to 9.02, a 0.08 of a year change.

Our results hint at the idea that the College expansion policy has had a smaller impact than the relaxation of the migration restrictions. However, given the highly non-linear nature of the model, comparing these two numbers is not necessarily the right comparison. In order to account for potential complementarities or substitutions arising from non-linearity, in the last column of the table we shut down both the migration reform and the college expansion. The result shows that it is still the case that the migration policy had a much larger impact, and that the college expansion is not nearly enough to compensate.

Finally, we look at how different levels of a subsidy to attending school would work in terms of increasing the incentive to invest in human capital. In Table 18 we illustrate three different levels of a subsidy to stay in school: 1.5 average monthly earnings, 1.5 average yearly earnings and 2.5 average yearly earnings. As the table shows, it is only when we get to the largest subsidy that we can undo the effects that the migration policy had on average schooling levels. As shown in Figure 8, the subsidy works mostly by encouraging people who would have stopped their schooling at middle school to continue on to high school.

9 Conclusion

The results in this paper point towards migration leading to less human capital accumulation, both for migrants and for those that stay behind. This is a consequence of a particular pattern of returns which reduces the overall expected return to school. We believe that this pattern is likely to arise whenever large low-skilled migration is present. Whether this reduced human capital accumulation should be concerning depends, in part, on whose point of view one takes. Absent market frictions, externalities, or other market failures, it is not clear that the outcome is not optimal from the perspective of the individuals making the decision. Sure, they may end up with less human capital, but they can potentially earn a lot more money by migrating. On the other hand, the receiving location may be better-off if immigrants had more human capital to begin with. Alternatively, if there is intergenerational transmission of human capital, the next generation may be worse-off depending on the relative importance of more parental human capital versus more resources. What kind of polices one may want to implement will depend on the, potentially conflicting, objectives of the different agents in the Economy.

References

BEINE, M., F. DOCQUIER, AND H. RAPOPORT (2001): "Brain Drain and Economic Growth: Theory and Evidence," *Journal of Development Economics*, 64, 275 – 289. 1

——— (2008): "Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers," *The Economic Journal*, 118, 631–652. 1

BLACK, D. A., T. G. MCKINNISH, AND S. G. SANDERS (2005): "Tight Labor Markets and the Demand for Education: Evidence from the Coal Boom and Bust," *Industrial and Labor Relations Review*, 59, 3–16. 1

BORJAS, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *The American Economic Review*, 77, 531–553. 1

——— (1995): "The Economic Benefits from Immigration," *Journal of Economic Perspectives*, 9, 3–22. 1

——— (2003): "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *The Quarterly Journal of Economics*, 118, 1335–1374. 1

BRANDT, L. AND C. HOLZ (2006): "Spatial Price Differences in China: Estimates and Implications," *Economic Development and Cultural Change*, 55, 43–86. 9

BRAUW, A. AND J. GILES (2017): "Migrant Opportunity and the Educational Attainment of Youth in Rural China," *The Journal of Human Resources*, 52, 272–311. 1

CARD, D. (2005): "Is the New Immigration Really so Bad?" *The Economic Journal*, 115, 300–323. 1

CHIQUIAR, D. AND G. HANSON (2005): "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113, 239–281. 1

CHISWICK, B. R. (1978): "The Effect of Americanization on the Earnings of Foreign-born Men," *Journal of Political Economy*, 86, 897–921. 1

CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010a): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931. 1

——— (2010b): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931. 6.2

DUSTMANN, C. AND A. GLITZ (2011): "Chapter 4 - Migration and Education," in *Handbook of The Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann, Elsevier, vol. 4 of *Handbook of the Economics of Education*, 327 – 439. 1, 2

FRENCH, E. AND C. TABER (2011): "Chapter 6 - Identification of Models of the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 4, 537 – 617. 6.2

HECKMAN, J. ., J. . STIXRUD, AND S. . URZUA (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482. 1

HECKMAN, J. J. (1974): "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42, 679–694. 1

HECKMAN, J. J., H. ICHIMURA, AND P. E. TODD (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64, 605–654. 1

HECKMAN, J. J. AND X. LI (2004): "Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000," *Pacific Economic Review*, 9, 155–171. 4

HECKMAN, J. J., L. LOCHNER, AND C. TABER (1999): "Human Capital Formation and General Equilibrium Treatment Effects: A Study of Tax and Tuition Policy," *Fiscal Studies*, 20, 25–40. 1

HECKMAN, J. J. AND S. MOSSO (2014): "The Economics of Human Development and Social Mobility," *Annual Review of Economics*, 6, 689–733. 1

HECKMAN, J. J. AND S. NAVARRO (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136, 341 – 396. 3.3, 6.2

HECKMAN, J. J. AND J. SCHEINKMAN (1987): "The Importance of Bundling in a Gorman-Lancaster Model of Earnings," *The Review of Economic Studies*, 54, 243–255. 1

HECKMAN, J. J. AND E. J. VYTLACIL (1999): "Local instrumental variables and latent variable models for identifying and bounding treatment effects," *Proceedings of the National Academy of Sciences*, 96, 4730–4734. 1

HENDRICKS, L. AND T. SCHOELLMAN (2017): "Human Capital and Development Accounting: New Evidence from Wage Gains at Migration*," *The Quarterly Journal of Economics*, 133, 665–700. 1

KENNAN AND WALKER (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79, 211–251. 5.2, 6.2

KNIGHT, J., S. LI, AND Q. DENG (2009): "Education and the Poverty Trap in Rural China: Setting the Trap," *Oxford Development Studies*, 37, 311–332. 1

LI, H., PRASHANTLOYALKA, S. ROZELLE, B. WU, AND J. XIE (2015): "Unequal Access to College in China: How Far Have Poor, Rural Students Been Left Behind?" *The China Quarterly*, 221, 185–207. 1

MACKINNON, J. G. (2019): "How cluster-robust inference is changing applied econometrics," *Canadian Journal of Economics*, 52, 851–881. 6, 7, 8, 9

MARIA, C. D. AND E. A. LAZAROVA (2012): "Migration, Human Capital Formation, and Growth: An Empirical Investigation," *World Development*, 40, 938 – 955. 1

MARIA, C. D. AND P. STRYSZOWSKI (2009): "Migration, Human Capital Accumulation and Economic Development," *Journal of Development Economics*, 90, 306 – 313. 1

MCKENZIE, D. AND H. RAPOPORT (2011): "Can migration reduce educational attainment? Evidence from Mexico," *Journal of Population Economics*, 24, 1331–1358. 4

MOUNTFORD, A. (1997): "Can a Brain Drain be Good for Growth in the Source Economy?" *Journal of Development Economics*, 53, 287 – 303. 1

NAVARRO, S. AND J. ZHOU (2017): "Identifying Agent's Information Sets: an Application to a Lifecycle Model of Schooling, Consumption and Labor Supply," *Review of Economic Dynamics*, 25, 58–92. 1, 3.3

STARK, O., C. HELMENSTEIN, AND A. PRSKAWETZ (1997): "A Brain Gain with a Brain Drain," *Economics Letters*, 55, 227 – 234. 1

STARK, O. AND Y. WANG (2002): "Inducing Human Capital Formation: Migration as a Substitute for Subsidies," *Journal of Public Economics*, 86, 29 – 46. 1

VIDAL, J.-P. (1998): "The Effect of Emigration on Human Capital Formation," *Journal of Population Economics*, 11, 589–600. 1

ZHOU, J. (2017): "Internal Migration with Social Networks in China," *Working paper*. 2, 7.2

10 Figures

Figure 1: College Offers

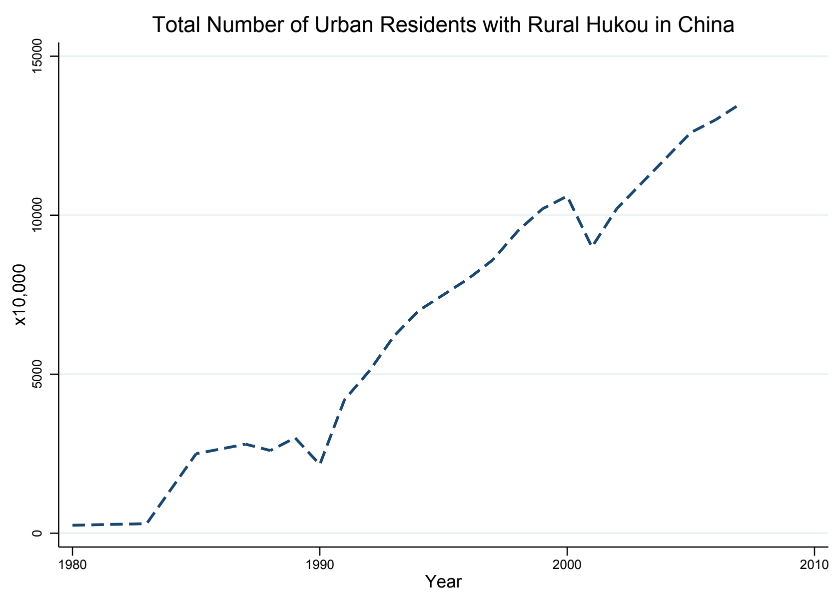
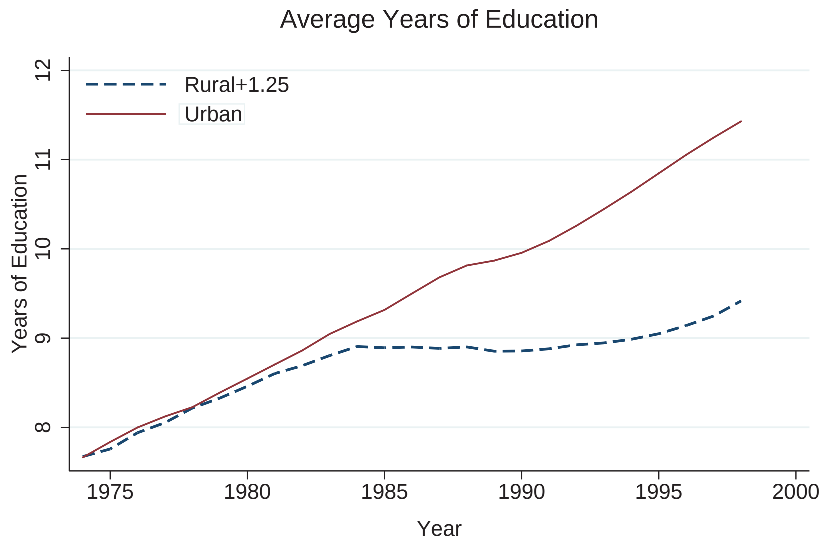
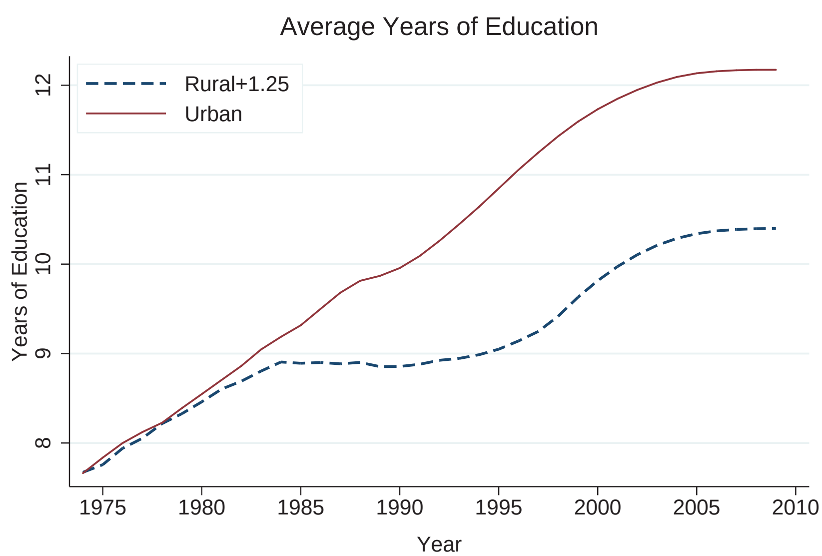


Figure 2: Number of Rural Migrants since 1980s



(a) 1974-1998



(b) 1974-2010

Figure 3: Comparison of Years of Education in China

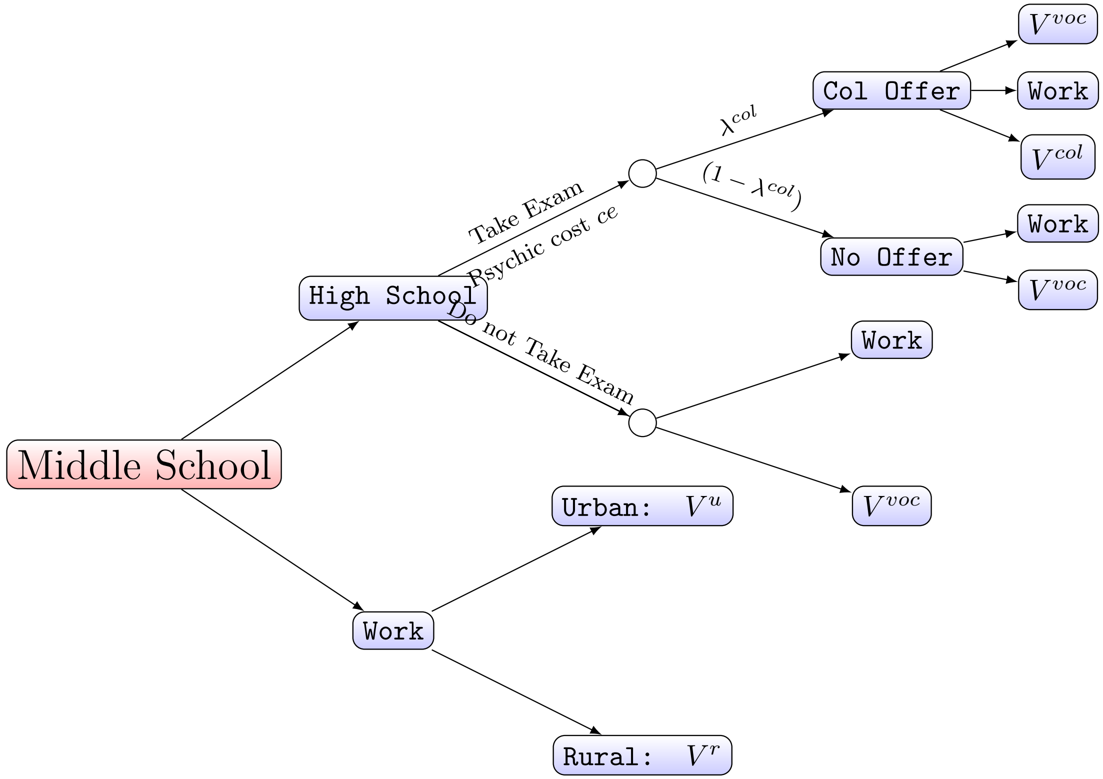
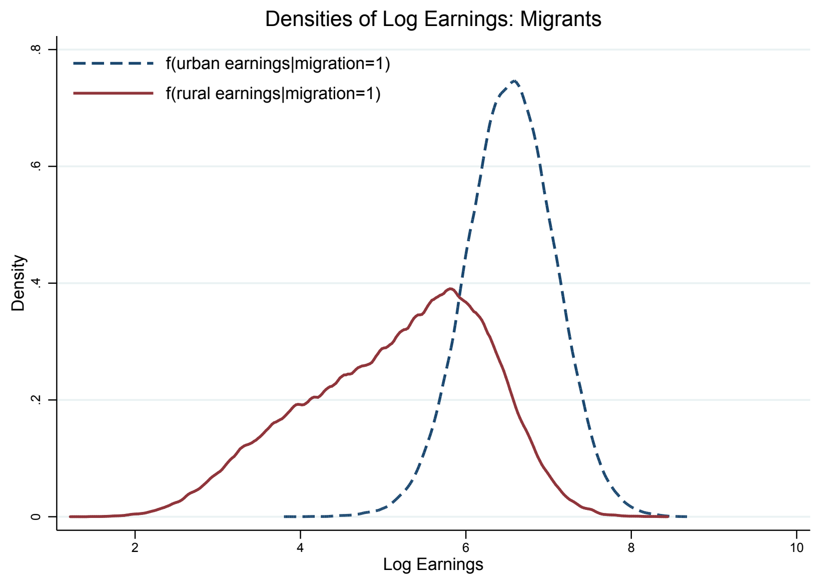
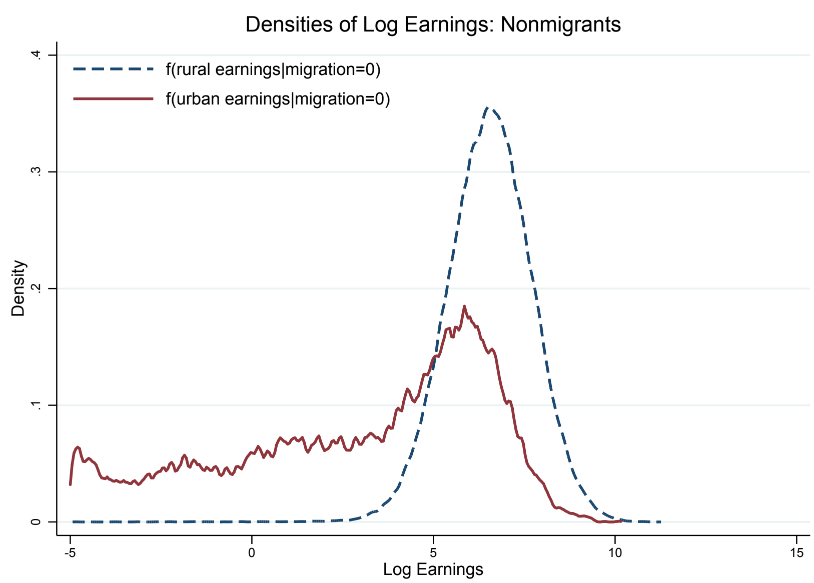


Figure 4: Model Timing: Education Decisions

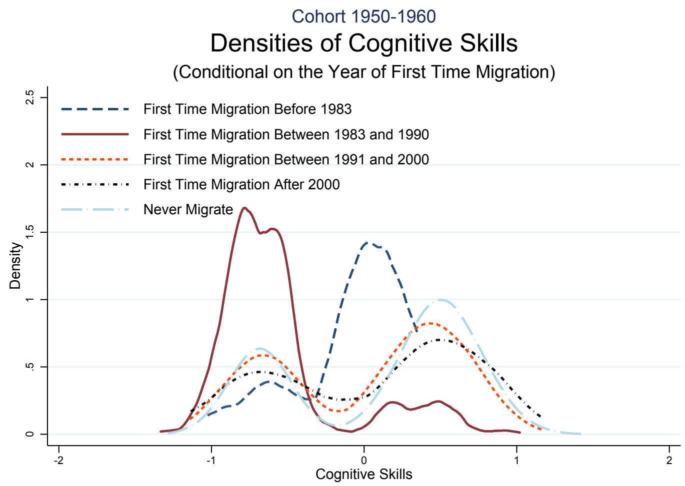


(a) Migrants

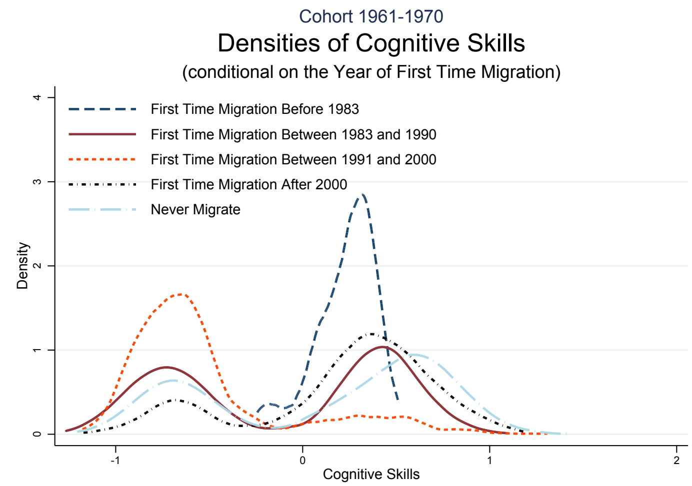


(b) Non-migrants

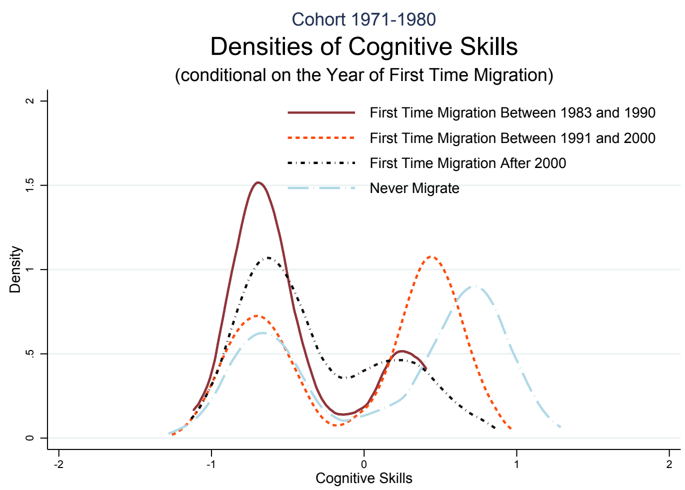
Figure 5: The Distribution of Log Earnings



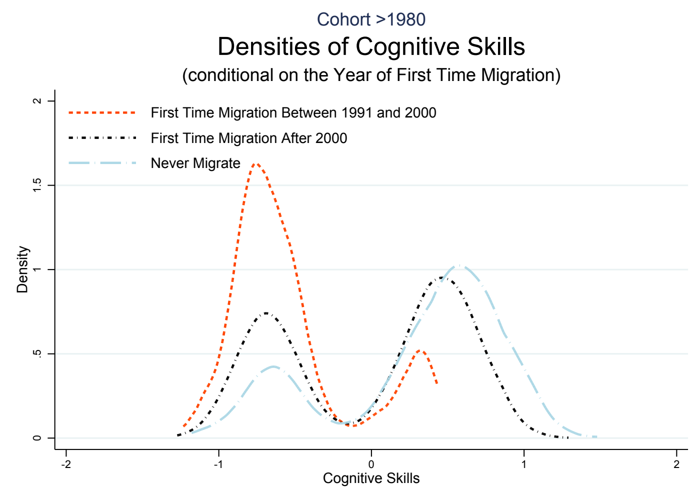
(a) Cohort 1950-1960



(b) Cohort 1961-1970

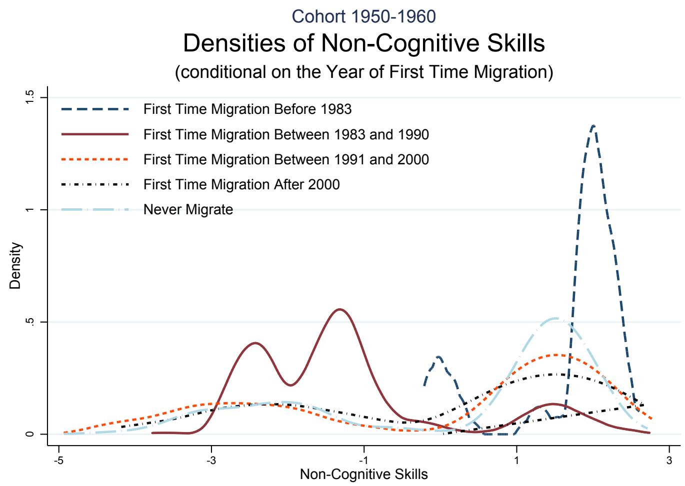


(c) Cohort 1971-1980

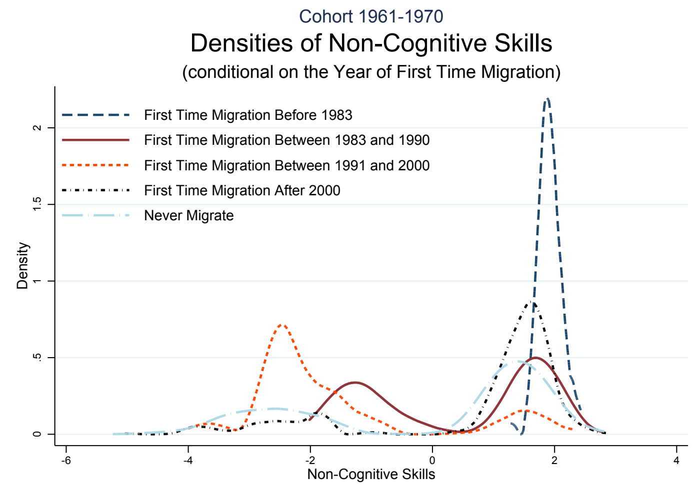


(d) Cohort > 1980

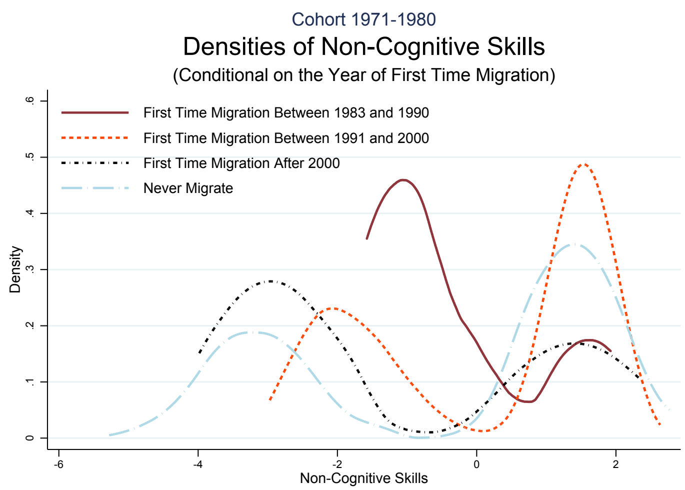
Figure 6: The Distribution of Cognitive Skills by First Time Migration Periods



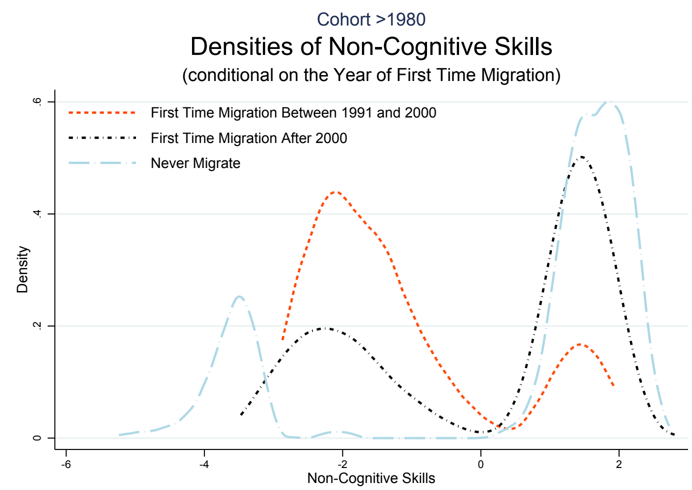
(a) Cohort 1950-1960



(b) Cohort 1961-1970



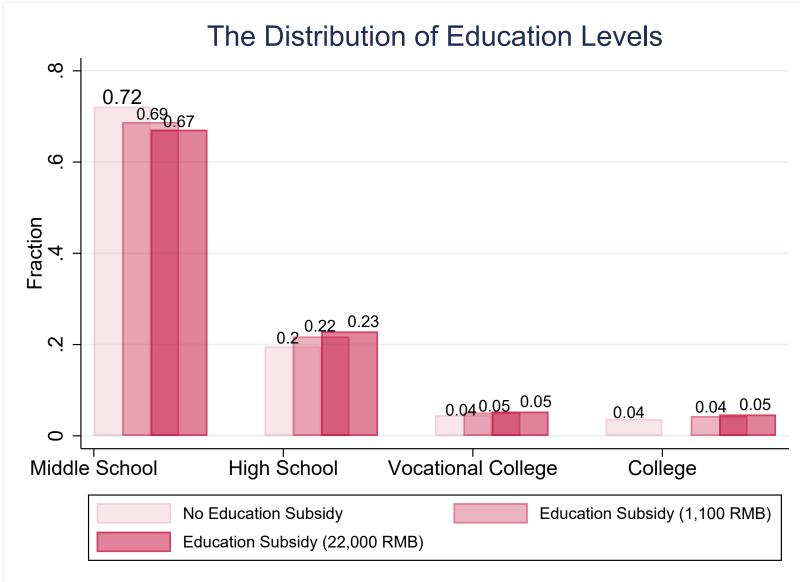
(c) Cohort 1971-1980



(d) Cohort > 1980

Figure 7: The Distribution of Non-Cognitive Skills by First Time Migration Periods

Figure 8: The Impacts of Education Subsidy Policies on Education Decisions



11 Tables

Table 1

Data Summary Statistics

Variables	Observations	Mean	Standard Deviation	Min	Max
		Urban			
Birth	4,251	1969.04	11.23	1951	1991
Years of Education	4,251	12.34	2.88	7	22
Class Performance	3,829	2.32	0.74	1	5
Take College Entrance Exam?	4,251	0.34	0.47	0	1
Can you concentrate to do something	3,037	1.61	0.66	1(Can)	4(Cannot)
Are you capable of making decisions	3,039	1.56	0.52	1(Always)	4(Do not)
Do you feel you cannot overcome difficulties?	3,039	1.59	0.50	1(Never)	4(Very Often)
Are you able to face problems?	3,039	1.53	0.61	1(Never)	4(Always)
Do you always lack confidence	3,039	1.41	0.47	1(Not at All)	4(Very Seriously)
		Rural			
Birth	8,292	1971.84	11.51	1951	1991
Years of Education	8,292	9.22	1.91	7	22
Class Performance	8,138	2.61	0.68	1	5
Take College Entrance Exam?	8,292	0.09	0.29	0	1
Can you concentrate to do something	4,873	1.46	0.63	1(Can)	4(Cannot)
Are you capable of making decisions	4,873	1.57	0.54	1(Always)	4(Do not)
Do you feel you cannot overcome difficulties?	4,872	1.56	0.50	1(Never)	4(Very Often)
Are you able to face problems?	4,874	1.51	0.56	1(Never)	4(Always)
Do you always lack confidence	4,872	1.42	0.48	1(Not at All)	4(Very Seriously)

Table 2

Data Summary Statistics: Rural Migrants and Rural Non-migrants

Variables	Observations	Mean	Standard Deviation	Min	Max
		Rural Non-migrants			
Log Monthly Earnings	4,543	6.51	1.11	-0.21	11.26
Birth	4,543	1966.82	10.83	1950	1991
Years of Education	4,543	9.01	1.70	7	19
Class Performance	4,543	2.59	0.68	1	5
Take College Entrance Exam?	4,543	0.09	0.28	0	1
Smoke	4,543	0.60	0.49	0	1
How many cigarettes usually smoke per day	4,543	8.58	10.46	0	91
		Rural Migrants			
Log Monthly Earnings	2,148	6.57	0.57	-0.08	10.19
Birth	2,148	1977.81	8.75	1950	1991
Years of Education	2,148	9.26	1.84	7	19
Class Performance	2,148	2.71	0.65	1	5
Take College Entrance Exam?	2,148	0.10	0.29	0	1
Smoke	2,148	0.50	0.50	0	1
How many cigarettes usually smoke per day	2,148	6.04	8.99	0	60

Table 3

The CHIPs Data Summary Statistics (For Auxiliary Model)

Variables	Observations	Mean	Standard Deviation	Min	Max
		2007-2009 Panel Data			
Birth	6,691	1970.35	11.42	1950	1991
Years of education	6,691	9.09	1.75	7	19
Education levels	6,691	1.43	0.80	1	4
1. Middle School	6,691	0.70			
2. High School	6,691	0.21			
3. Vocation College	6,691	0.05			
4. College or Above	6,691	0.04			
Log monthly earnings					
in 2007	6,691	6.53	0.97	-0.21	11.26
in 2008	6,691	6.52	0.96	-0.26	10.43
in 2009	6,691	6.59	1.00	-0.26	10.67
Migration status					
in 2007	6,691	0.32	0.47	0	1
in 2008	6,691	0.31	0.46	0	1
in 2009	6,691	0.34	0.47	0	1
Year of first time migration	3,275	1999.33	6.80	1967	2008
Ever migrate	6,691	0.58	0.49	0	1
Class performance	6,691	2.63	0.68	1	5
Take college entrance exam?	6,691	0.09	0.28	0	1
Smoke	6,691	0.57	0.50	0	1
How many cigarettes usually smoke per day	6,691	7.76	10.08	0	91

Table 4

CHIPS Data Cognitive and Non-Cognitive Skills Summary Statistics

Variables	Mean	Standard Deviation	Min	Max
Birth	1960.80	7.78	1950	1989
Years of education	7.87	2.32	1	19
Age at first round Test	46.20	7.78	18	57
Class performance	2.61	0.70	1	5
Take college entrance exam?	0.07	0.25	0	1
Non-Cognitive skill measures				
Can you concentrate to do something?				
2007	1.45	0.69	1 (Can)	4 (Cannot)
2008	1.50	0.72	1 (Can)	4 (Cannot)
2009	1.58	0.74	1 (Can)	4 (Cannot)
Are you capable of making decisions?				
2007	1.59	0.58	1 (Always)	4 (Do not)
2008	1.64	0.61	1 (Always)	4 (Do not)
2009	1.57	0.56	1 (Always)	4 (Do not)
Do you feel you cannot overcome difficulties?				
2007	1.50	0.61	1 (Never)	4 (Very Often)
2008	1.52	0.63	1 (Never)	4 (Very Often)
2009	1.59	0.64	1 (Never)	4 (Very Often)
Are you able to face problems?				
2007	1.50	0.61	1 (Never)	4 (Always)
2008	1.52	0.63	1 (Never)	4 (Always)
2009	1.59	0.64	1 (Never)	4 (Always)
Do you always lack of confidence?				
2007	1.46	0.54	1 (Not at All)	4 (Very Seriously)
2008	1.45	0.54	1 (Not at All)	4 (Very Seriously)
Smoke	0.69	0.46	0	1
How many cigarettes usually smoke per day	9.91	10.83	0	91
Observations	3,742			

Table 5

CFPS Data Cognitive and Non-Cognitive Skills Summary Statistics

Variables	Mean	Standard Deviation	Min	Max
Birth	1966.99	10.70	1950	1991
Years of education	8.32	4.01	0	19
Age at first round test	43.01	10.70	19	60
Standardized cognitive skills				
Math in 2010	0.29	0.87	-1.47	2.15
Language in 2010	0.28	0.84	-1.56	1.60
Immediately word recall in 2012	0.11	0.91	-2.26	2.82
Delayed word recall in 2012	0.10	0.93	-1.53	3.09
Math in 2014	0.27	0.89	-1.42	2.24
Language in 2014	0.27	0.87	-1.44	1.63
Get on well with others	4.07	0.84	1 (Very hard)	5 (Very easy)
Feel upset and cannot remain calm	4.56	0.79	1 (Almost everyday)	5 (Never)
Feel everything is difficult	4.46	0.86	1 (Almost everyday)	5 (Never)
Can you concentrate to do something	1.55	0.74	1 (Almost never)	4 (Most of the time)
Smoke	0.59	0.49	0	1
How many cigarettes usually smoke per day	11.03	17.45	0	400
How much alcohol do you drink last month in 2010	21.33	52.03	0	770
How much alcohol do you drink last week in 2012	11.42	11.69	0	80
How much alcohol do you drink last week in 2014	11.32	11.46	0	100
Observations	6,033			

Table 6

Return to Education
(Rural Males)

	Log Monthly Earnings				
	(1)	(2)	(3)	(4)	(5)
Age	0.1238***	0.1229***	0.1231***	0.1220***	0.1222***
	(0.0126)	(0.0125)	(0.0124)	(0.0121)	(0.0123)
Age^2	-0.0017***	-0.0017***	-0.0017***	-0.0017***	-0.0017***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Years of Education	0.0559***	0.0564***	0.0501***	0.0505***	0.0393+
	(0.0100)	(0.0100)	(0.0132)	(0.0132)	(0.0212)
Migration	0.3678***	0.3654***	0.3800**	0.3777**	0.4817**
	(0.0988)	(0.1001)	(0.1061)	(0.1072)	(0.1353)
Years of Education × Migration	-0.0358**	-0.0356**	-0.0371**	-0.0369**	-0.0491*
	(0.0112)	(0.0111)	(0.0110)	(0.0110)	(0.0195)
Vocation College					0.0251
					(0.0954)
College					0.1927
					(0.1613)
Years of Education × Migration × Vocation					0.0095
					(0.0095)
Years of Education × Migration × College				0.0039	
					(0.0117)
Constant	3.9822***	3.9857***	4.0796***	4.0908***	4.1744***
	(0.2458)	(0.2461)	(0.2095)	(0.2083)	(0.2166)
Non-Cognitive Skill Measures	No	Yes	No	Yes	Yes
Cognitive Skill Measures	No	No	Yes	Yes	Yes

Standard errors in parentheses, clustered at province × year level using the wild bootstrap procedure in MacKinnon (2019)
$^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Table 7

Migration and Education Decisions: Pretrend Analysis

	Years of Education			
	(1)	(2)	(3)	(4)
Year 1975 × Rural	-0.0916	-0.0890	-0.0802	-0.0810
	(0.2448)	(0.3321)	(0.3093)	(0.3095)
Year 1976 × Rural	-0.0734	-0.0708	-0.0621	-0.0652
	(0.2570)	(0.3375)	(0.3132)	(0.3132)
Year 1977 × Rural	-0.0887	-0.0819	-0.0655	-0.0691
	(0.2731)	(0.3480)	(0.3250)	(0.3255)
Year 1978 × Rural	-0.0313	-0.0223	0.0024	-0.0016
	(0.2852)	(0.3545)	(0.3338)	(0.3345)
Year 1979 × Rural	-0.0857	-0.0757	-0.0414	-0.0467
	(0.2742)	(0.3431)	(0.3239)	(0.3242)
Year 1980 × Rural	-0.1125	-0.1013	-0.0585	-0.0653
	(0.2728)	(0.3430)	(0.3259)	(0.3262)
Year 1981 × Rural	-0.1289	-0.1192	-0.0663	-0.0757
	(0.2740)	(0.3386)	(0.3247)	(0.3238)
Year 1982 × Rural	-0.2003	-0.1904	-0.1234	-0.1357
	(0.2535)	(0.3143)	(0.3015)	(0.3004)
Rural	-1.1091***	-1.2287***	-1.3130***	-1.2990***
	(0.1697)	(0.2300)	(0.2129)	(0.2157)
Constant	7.7458***	7.9107***	7.8230***	7.8808***
	(0.1153)	(0.2417)	(0.2043)	(0.2435)
Year Dummies	Yes	Yes	Yes	Yes
Province Dummies	Yes	Yes	Yes	Yes
Cognitive Skill Measures	No	Yes	No	Yes
Non-cognitive Skill Measures	No	No	Yes	Yes

Standard errors in parentheses, clustered at province × year level using the wild bootstrap procedure in MacKinnon (2019)

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8

The Effect of Relaxing Migration Restrictions on Education
(1975-1989)

	Years of Education					
	(1)	(2)	(3)	(4)	(5)	(6)
Rural	-1.3223***	-1.2325***	-1.2332***	-1.1641***	-1.2922***	-1.2271***
	(0.2351)	(0.2141)	(0.2191)	(0.1972)	(0.2048)	(0.1828)
$>$ 1982	1.0486***	1.0640***	1.7919***	1.8014***	1.8765***	1.8849***
	(0.1624)	(0.1727)	(0.0574)	(0.0496)	(0.0673)	(0.0589)
$>$ 1982 $\times$ Rural	-0.4031***	-0.4135***	-0.4134***	-0.4020***	-0.3312***	-0.3230***
	(0.1283)	(0.1369)	(0.1412)	(0.1439)	(0.1199)	(0.1227)
Constant	8.3055***	8.5276***	7.9308***	8.7000***	8.1661***	8.7261***
	(0.2274)	(0.2183)	(0.1913)	(0.5028)	(0.2453)	(0.5297)
Cognitive Skill Measures	No	No	No	No	Yes	Yes
Non-Cognitive Skill Measures	No	No	No	Yes	No	Yes
Year Dummies	No	No	Yes	Yes	Yes	Yes
Province Dummies	No	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, clustered at province $\times$ year level using the wild bootstrap procedure in MacKinnon (2019).
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
$>$ 1982 is an indicator for whether the individual finishes his schooling after the migration reform 1982.

Table 9

Educational Attainment for Likely Migrants

	Years of Education		
	$\pi = 1(P_{mig} > 0.5)$ $\pi = 0(P_{mig} \leq 0.5)$	$\pi = 1(P_{mig} > 0.7)$ $\pi = 0(P_{mig} \leq 0.3)$	$\pi = 1(P_{mig} > 0.85)$ $\pi = 0(P_{mig} \leq 0.15)$
π	-0.106	0.201	0.0500
	(0.109)	(0.176)	(0.316)
$>$ 1982	0.185*	0.400**	0.423
	(0.0797)	(0.137)	(0.377)
$\pi \times$ $>$1982	-0.125	-0.606**	-0.892*
	(0.138)	(0.217)	(0.426)
Constant	8.681***	8.583***	9.000***
	(0.0585)	(0.0959)	(0.259)
Observations	2175	877	394

We run a probit for the probability that an individual will migrate using only pre-1982 data. We then use this pre-1982 model to predict how likely an individual would have been to migrate post-1982 as if the policy had not been in place to generate π.
$>$ 1982 is an indicator for whether the individual finishes his schooling after the migration reform 1982.
Standard errors in parentheses,clustered at province $\times$ year level using the wild bootstrap procedure in MacKinnon (2019)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10

Model Fit: Log Earnings Regression

2007 Log Earnings Regression			Log Earning Regression for Never Migrants			Migrant's Log Earnings (with Full History)		
	Data	Model		Data	Model		Data	Model
Years of Education	0.0406	0.0458	Years of Education	0.0381	0.0422	Years of Education	0.0191	0.0095
	(0.0100)			(0.0159)			(0.0130)	
Age	0.1220	0.1259	Experience of Rural	0.0730	0.0943	Experience of Rural	0.0081	0.0082
	(0.0073)			(0.0089)			(0.0063)	
Age^2	-0.0017	-0.0018	Experience of Rural^2	-0.0016	-0.0023	Experience of Rural^2	-0.0001	-0.0019
	(0.0001)			(0.0002)			(0.0002)	
Migration	0.4081	0.3722	Cog 1	-0.0467	-0.0093	Experience of Urban	0.0532	0.0288
	(0.1309)			(0.0320)			(0.0088)	
Years of Education $\times$ Migration	-0.0400	-0.0447	Cog 2	0.1397	-0.0026	Experience of Urban^2	-0.0013	-0.0001
	(0.0139)			(0.0831)			(0.0004)	
Cog 1	-0.0138	0.0135	Non cog 1	0.0074	-0.0003	Cog 1	-0.0127	-0.0137
	(0.0178)			(0.0488)			(0.0236)	
Cog 2	0.0447	-0.0193	Non cog 2	-0.0002	0.0000	Cog 2	0.0343	-0.0003
	(0.0479)			(0.0022)			(0.0673)	
Non cog 1	0.0122	0.0251	College	0.3217	0.2738	Non cog 1	0.0425	0.0489
	(0.0273)			(0.1363)			(0.0339)	
Non cog 2	0.0007	-0.0015	Constant	5.5918	5.4386	Non cog 2	-0.0026	-0.0028
	(0.0013)			(0.2168)			(0.0020)	
College	0.1843	0.2101	Mean of Log Earnings	6.4227	6.4166	College	0.2273	0.2128
	(0.0754)						(0.0995)	
Constant	4.1715	4.1385	Variance of Log Earnings	1.2672	1.2055	Constant	6.0915	6.1540
	(0.1703)						(0.1539)	
Mean of Log Earnings	6.5306	6.5484				Mean of Log Earnings	6.5695	6.5003
Variance of Log Earnings	0.9442	1.0500				Variance of Log Earnings	0.3334	0.2754

Table 11

Model Fit: Migration and Education Moments

	Data	Model
Year of First Time Migration		
Cohort:1950-1960	1993.48	1988.91
	(10.07)	
Cohort:1961-1970	1996.05	1990.79
	(7.83)	
Cohort:1971-1980	1999.30	1994.92
	(4.76)	
Cohort: $>$1980	2004.29	2002.73
	(2.88)	
Fraction Ever Migrated		
Cohort:1950-1960	0.25	0.28
Cohort:1961-1970	0.51	0.49
Cohort:1971-1980	0.71	0.74
Cohort: $>$1980	0.84	0.80
Average Years of Education	9.09	9.11
	(1.75)	
Fraction Attending High School or Above	0.29	0.26
Conditional on Attending High School		
Fraction Attending Vocational College	0.21	0.20
Fraction Attending College	0.18	0.16
Fraction Taking College Entrance Exam	0.35	0.34

Table 12

Migration Regression

Return Migration Regression I			Return Migration Regression II		
	Data	Model		Data	Model
Years of Education	0.0027	0.0053	Years of Education	-0.0078	-0.0066
	(0.0059)			(0.0031)	
Age	-0.0272	-0.0300	Age	0.0055	-0.0145
	(0.0064)			(0.0029)	
Age^2	0.0005	0.0004	Age^2	-0.0001	0.0002
	(0.0001)			(0.0000)	
Cognitive Skill 1	0.0136	0.0028	Cognitive Skill 1	-0.0051	0.0008
	(0.0135)			(0.0071)	
Cognitive Skill 2	-0.0078	-0.0095	Cognitive Skill 2	0.0095	-0.0020
	(0.0357)			(0.0188)	
Noncognitive Skill 1	0.0234	-0.0063	Noncognitive Skill 1	0.0130	0.0007
	(0.0195)			(0.0108)	
Noncognitive Skill 2	-0.0017	0.0000	Noncognitive Skill 2	0.0004	-0.0001
	(0.0011)			(0.0005)	
Constant	0.4381	0.5387	Location in 2007	-0.9887	-0.9456
	(0.1222)			(0.0019)	
			Constant	0.9494	1.2841
				(0.0635)	

Migration Regression I			Migration Regression II		
	Data	Model		Data	Model
Age	-0.0020	-0.0480	Years of Education	0.0007	-0.0069
	(0.0061)			(0.0081)	
Age^2	-0.0002	0.0005	Cognitive Skill 1	-0.0027	0.0055
	(0.0001)			(0.0155)	
Years of Education	-0.0058	0.0520	Cognitive Skill 2	0.0721	0.0090
	(0.0071)			(0.0674)	
Cohort(1961-1970)	0.0567	-0.5727	Noncognitive Skill 1	-0.0066	-0.0296
	(0.0872)			(0.0252)	
Cohort(1971-1980)	-0.0656	0.0562	Noncognitive Skill 2	0.0009	0.0012
	(0.0925)			(0.0010)	
Cohort($>$1980)	0.1353	0.5965	Constant	0.0308	0.2161
	(0.0990)			(0.0826)	
Years of Education$\times$Cohort(1961-1970)	-0.0106	0.0644			
	(0.0093)				
Years of Education$\times$Cohort(1971-1980)	0.0161	-0.0046			
	(0.0091)				
Years of Education$\times$Cohort(>1980)	-0.0013	-0.0723			
	(0.0088)				
Cognitive Skill 1	0.0159	-0.0010			
	(0.0078)				
Cognitive Skill 2	0.0271	-0.0043			
	(0.0209)				
Noncognitive Skill 1	0.0303	-0.0017			
	(0.0120)				
Noncognitive Skill 2	-0.0009	0.0002			
	(0.0006)				
Constant	0.5908	0.8358			
	(0.1447)				

Table 13

Structural Model Estimation Parameters (Part I)

Earning Equations		
	Rural	Urban
Years of Education	0.0360	0.0158
	(0.0078)	(0.0084)
Rural Work Experience	0.0907	0.0098
	(0.0111)	(0.0030)
Urban Work Experience	0.7480	0.0258
	(0.0009)	(0.0466)
Rural Work Experience^2	-0.0023	-0.0021
	(0.0031)	(0.0002)
Urban Work Experience^2	-0.2689	0.0000
	(0.0000)	(0.0143)
Cognitive Skill	-0.0084	0.0283
	(0.0000)	(0.0099)
Non-cognitive Skill	-0.0080	-0.0481
	(0.0000)	(0.0001)
College	0.3379	0.5978
	(0.0038)	(0.0006)
Constant	5.5363	6.0541
	(0.0131)	(0.0027)
Variance of shocks	1.0767	0.2392
	(0.0833)	(0.0219)

Table 14

Structural Model Estimation Parameters (Part II)

Migration costs		Psychic Value of Living in Rural	
Age	0.0839	Age	-11.5918
	(0.0022)		(0.0003)
Age^2	0.0020	Age^2	0.2152
	(0.0006)		(0.0000)
Year Trend	-0.3482	Cognitive Skill	29.8322
	(0.0028)		(0.0029)
Cognitive Skill	0.1633	Non-Cognitive Skill	0.3243
	(0.0013)		(0.0002)
Non-Cognitive Skill	-4.1509	Constant	2.3947
	(0.0139)		(0.0001)
Whether After 1982		Terminal Values	
Year Trend	-0.0567	Rural Work Experience	0.0639
	(0.0007)		(0.0000)
Constant	-0.2479	Urban Work Experience	5.4146
	(0.0017)		(0.0001)
Whether After 1990		Rural Experience^2	11.4459
Year Trend	-0.1141		(0.0000)
	(0.0015)	Urban Work Experience^2	0.3718
Constant	-0.0029		(0.0348)
	(0.0010)	Years of Education	0.4572
Constant	29.8249		(0.0001)
	(0.0663)	Cognitive Skill	-0.6711
			(0.0527)
		Non-Cognitive Skill	-0.0369
			(0.0451)
		Constant	-0.0016
			(0.0049)

Table 15

Structural Model Estimation Parameters (Part III)

Education Utility	
High School Level	
Cognitive Skill	0.1949
	(0.0010)
Noncognitive Skill	-0.1167
	(0.0016)
Year Trend	-0.0079
	(0.0008)
$1_{1965<year<1977}$	0.0820
	(0.0004)
Constant	-273.8736
	(0.1268)
Some College Level	
Cognitive Skill	0.5720
	(0.0003)
Noncognitive Skill	-0.3171
	(0.0008)
Year Trend	-2.1580
	(0.0001)
$1_{1965<year<1977}$	0.0023
	(0.0001)
Constant	-86.3824
	(0.0391)
College Level	
Cognitive Skill	6.3682
	(0.0016)
Noncognitive Skill	-0.0018
	(0.0001)
Year Trend	-0.0080
	(0.0002)
$1_{1965<year<1977}$	-0.0005
	(0.0002)
Constant	283.5237
	(0.0651)

Probability of Getting College Offer	
Cognitive Skill	-7.1281
	(0.0005)
Non cognitive Skill	2.4212
	(0.0000)
Year Trend	-0.0094
	(0.0000)
After 1999	
Year Trend	0.0077
	(0.0000)
Constant	-0.0053
	(0.0000)
Constant	-4.4308
	(0.0006)

Psychic Cost of Taking College Entrance Exam	
Cognitive Skill	-0.0339
	(0.0029)
Non cognitive Skill	20.9558
	(0.0534)
Constant	-426.2201
	(9.6534)

Table 16: Counterfactual Analysis

	Model	No Migration Policy	No College Expansion	Neither
Average Years of Education	9.11	9.34	9.02	9.25
Fraction Attending High School or Above	0.26	0.30	0.23	0.28
Conditional on Attending High School				
Fraction Attending Some College	0.20	0.31	0.22	0.33
Fraction Attending College	0.16	0.12	0.11	0.08
Fraction Taking College Entrance Exam	0.34	0.26	0.26	0.19
Fraction of Ever Migration				
Cohort:1950-1960	0.28	0.23	0.28	0.23
Cohort:1961-1970	0.49	0.27	0.49	0.27
Cohort:1971-1980	0.74	0.35	0.74	0.35
Cohort: $>$1980	0.80	0.44	0.85	0.48
Mean of Log Earnings in 2007	6.55	6.51	6.54	6.50
Variance of Log Earning in 2007	1.05	1.17	1.05	1.17
Mean of Log Earnings for Never Migrants in 2007	6.42	6.47	6.43	6.47
Variance of Log Earnings for Never Migrants in 2007	1.21	1.18	1.20	1.18

Table 17: Education Transitions

	No Migration Policy			
Benchmark	Middle School	High School	Some College	College
Middle School	93.63%	6.35%	0.02%	0.00%
High School	1.54%	75.24%	23.22%	0.00%
Some College	0.00%	0.00%	100.00%	0.00%
College	0.54%	0.00%	0.00%	99.46%

Table 18: The Impacts of Education Subsidy Policies

	Education Subsidy		
	1,100 RMB	8,100 RMB	22,000 RMB
Increase in Avg. Education	0.14	0.17	0.24

1. 1,100 RMB is approx migrants' 1.5 monthly earnings
2. 8,100 RMB is approx migrants' 1.5 annual earnings
3. 22,000 RMB is approx migrants' 2.5 annual earnings

Appendix
Human Capital and Migration: a Cautionary Tale

Salvador Navarro, and Jin Zhou

2020/07/15, 11:09am

Contents

List of Tables

A Estimates

Table A1

Return to Education

(Rural Females)

	Log Earnings				
Age	0.0771***	0.0765***	0.0768***	0.0763***	0.0763***
	(0.0129)	(0.0133)	(0.0129)	(0.0134)	(0.0134)
Age^2	-0.0010***	-0.0010***	-0.0010***	-0.0010***	-0.0010***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Years of Education	0.0519***	0.0478***	0.0458***	0.0426***	0.0421***
	(0.0127)	(0.0117)	(0.0108)	(0.0100)	(0.0103)
Migration	0.4337***	0.4153***	0.4572***	0.4379***	0.4792***
	(0.0824)	(0.0860)	(0.0859)	(0.0895)	(0.1008)
Years of Education × Migration	-0.0203	-0.0177	-0.0228	-0.0202	-0.0254^{+}
	(0.0138)	(0.0128)	(0.0142)	(0.0133)	(0.0144)
Vocation College					0.0915
					(0.1591)
College					0.0635
					(0.2821)
Years of Education × Migration × Vocation					0.0107
					(0.0131)
Years of Education × Migration × College					0.0083
					(0.0177)
Constant	4.3690***	4.8372***	4.5392***	4.9536***	4.9545***
	(0.2292)	(0.2187)	(0.2031)	(0.2397)	(0.2422)
Non-Cognitive Skill Measures	No	Yes	No	Yes	Yes
Cognitive Skill Measures	No	No	Yes	Yes	Yes

Standard errors in parentheses, clustered in the province level using the wild bootstrap procedure in MacKinnon (2019)

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2

Migration and Education Decisions: Pretrend Analysis

(Rural Females)

	Years of Education			
	(1)	(2)	(3)	(4)
Year 1977× Rural	-0.196	-0.254	-0.341	-0.380
	(0.301)	(0.300)	(0.282)	(0.282)
Year 1978× Rural	0.311	0.306	0.210	0.194
	(0.299)	(0.298)	(0.280)	(0.280)
Year 1979× Rural	0.110	0.114	-0.010	-0.002
	(0.306)	(0.305)	(0.287)	(0.286)
Year 1980× Rural	0.309	0.281	0.412	0.392
	(0.297)	(0.295)	(0.278)	(0.278)
Year 1981× Rural	-0.095	-0.101	-0.109	-0.112
	(0.304)	(0.303)	(0.285)	(0.284)
Year 1977	-0.148	-0.094	-0.035	0.003
	(0.248)	(0.247)	(0.231)	(0.231)
Year 1978	-0.408	-0.406	-0.260	-0.248
	(0.245)	(0.244)	(0.229)	(0.228)
Year 1979	-0.437	-0.428	-0.347	-0.344
	(0.252)	(0.251)	(0.235)	(0.235)
Year 1980	-0.340	-0.307	-0.407	-0.383
	(0.243)	(0.242)	(0.228)	(0.228)
Year 1981	0.188	0.183	0.165	0.162
	(0.248)	(0.246)	(0.231)	(0.231)
Rural	-2.833***	-2.800***	-2.262***	-2.266***
	(0.219)	(0.219)	(0.209)	(0.209)
Constant	10.560***	11.517***	15.224***	15.698***
	(0.180)	(0.256)	(0.300)	(0.334)
Year Dummies	Yes	Yes	Yes	Yes
Province Dummies	Yes	Yes	Yes	Yes
Cognitive Skill Measures	No	Yes	No	Yes
Non-cognitive Skill Measures	No	No	Yes	Yes
Observations	2,337	2,337	2,319	2,319

Standard errors in parentheses, clustered at province × year level using the wild bootstrap procedure in MacKinnon (2019)

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3

The Effect of Relaxing Migration Restrictions on Education

(Rural Females)

	Years of Education					
	(1)	(2)	(3)	(4)	(5)	(6)
Rural	-1.563***	-1.529***	-1.542***	-1.505***	-1.331***	-1.288***
	(0.090)	(0.084)	(0.084)	(0.080)	(0.089)	(0.087)
>1982	1.105***	1.123***	2.115***	2.119***	2.238***	2.241***
	(0.078)	(0.057)	(0.103)	(0.104)	(0.099)	(0.100)
>1982× Rural	-0.218*	-0.227**	-0.222**	-0.230**	-0.158	-0.163
	(0.090)	(0.081)	(0.079)	(0.079)	(0.087)	(0.089)
Constant	7.991***	8.413***	7.715***	7.861***	7.791***	7.897***
	(0.078)	(0.088)	(0.112)	(0.208)	(0.117)	(0.244)
Cognitive Skill Measures	No	No	No	No	Yes	Yes
Non-Cognitive Skill Measures	No	No	No	Yes	No	Yes
Year Dummies	No	No	Yes	Yes	Yes	Yes
Province Dummies	No	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, clustered at province × year level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

> 1982 is an indicator for whether the individual finishes his schooling after the migration reform in 1982

Table A4

The Effect of Relaxing Migration Restrictions on Education

(1975-2009)

(Rural Males)

	Years of Education					
	(1)	(2)	(3)	(4)	(5)	(6)
Rural	-1.322***	-1.311***	-1.313***	-0.971***	-1.352***	-1.012***
	(0.0846)	(0.0815)	(0.0817)	(0.0827)	(0.0799)	(0.0811)
>1982	2.660***	2.662***	4.273***	4.152***	4.351***	4.229***
	(0.0787)	(0.0728)	(0.101)	(0.111)	(0.0995)	(0.110)
>1982×Rural	-1.360***	-1.356***	-1.415***	-1.296***	-1.280***	-1.166***
	(0.0846)	(0.0808)	(0.0803)	(0.0828)	(0.0786)	(0.0799)
Constant	8.305***	8.461***	7.877***	11.260***	8.457***	11.530***
	(0.0787)	(0.0729)	(0.103)	(0.544)	(0.149)	(0.606)
Cognitive Skill Measures	No	No	No	No	Yes	Yes
Non-Cognitive Skill Measures	No	No	No	Yes	No	Yes
Year Dummies	No	No	Yes	Yes	Yes	Yes
Province Dummies	No	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, clustered at province × year level using the wild bootstrap procedure in MacKinnon (2019).

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

> 1982 is an indicator for whether the individual finishes his schooling after the migration reform 1982.

Table A5

Cognitive Skill Measures (CFPS)

	Math 2010		Word 2010	
	Data	Model	Data	Model
Years of Education at the Test Time	0.1775	0.1710	0.1368	0.1354
	(0.0016)		(0.0020)	
Age at the Test Time	-0.0029	-0.0026	-0.0076	-0.0085
	(0.0006)		(0.0008)	
Constant	-1.0653	-1.0751	-0.5360	-0.5343
	(0.0325)		(0.0408)	
Residual^2	0.2346	0.3963	0.3694	0.3044
	Delayed Record		Immediate Recall	
	Data	Model	Data	Model
Years of Education at the Test Time	0.0621	0.0583	0.0638	0.0657
	(0.0028)		(0.0027)	
Age at the Test Time	-0.0254	-0.0285	-0.0222	-0.0210
	(0.0010)		(0.0010)	
Constant	0.7275	0.7100	0.5783	0.5827
	(0.0578)		(0.0569)	
Residual^2	0.6964	0.6681	0.6734	0.6034
	Math 2014		Word 2014	
	Data	Model	Data	Model
Years of Education at the Test Time	0.1884	0.1769	0.1384	0.1286
	(0.0016)		(0.0021)	
Age at the Test Time	-0.0010	-0.0016	-0.0113	-0.0114
	(0.0006)		(0.0008)	
Constant	-1.2535	-1.2275	-0.3528	-0.3513
	(0.0316)		(0.0450)	
Residual^2	0.2220	0.1068	0.3960	0.3926

Table A6

Model Fit: Non-cognitive Skill Measures CHIP

	Deal with Things Decisively (2007-2009)					
	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.0363	0.0363	0.0365	0.0325	0.0362	0.0367
	(0.0007)		(0.0007)		(0.0007)	
Age at the Test Time	0.0147	0.0149	0.0147	0.0089	0.0147	0.0149
	(0.0001)		(0.0001)		(0.0001)	
D=2	-0.9676	-0.6099	-0.9675	-0.9644	-0.9628	-0.9703
	(0.0042)		(0.0042)		(0.0042)	
D=3	-0.9208	-0.9196	-0.9472	-0.9364	-0.9457	-0.9556
	(0.0110)		(0.0105)		(0.0091)	
D=4	-0.9529	-0.6662	-0.9513	-1.0154	-0.9776	-0.9458
	(0.0282)		(0.0475)		(0.0289)	
Residual^2	0.0154	0.0123	0.0155	0.0158	0.0154	0.0158

	Impossible to Overcome Difficulties (2007-2009)					
	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.0364	0.0420	0.0363	0.0364	0.0363	0.0364
	(0.0007)		(0.0007)		(0.0007)	
Age at the Test Time	0.0147	0.0134	0.0147	0.0151	0.0147	0.0142
	(0.0001)		(0.0001)		(0.0001)	
D=2	-0.9631	-0.9339	-0.9641	-0.9235	-0.9620	-0.9690
	(0.0041)		(0.0041)		(0.0042)	
D=3	-0.9495	-0.9226	-0.9381	-0.9315	-0.9527	-0.8497
	(0.0136)		(0.0135)		(0.0122)	
D=4	-0.9494	-0.8903	-0.9903	-0.9468	-1.0070	-0.8197
	(0.0379)		(0.0419)		(0.0349)	
Residual^2	0.0154	0.0160	0.0154	0.0157	0.0154	0.0160

	Escape from the Difficulties (2007-2009)					
	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.0364	0.0371	0.0365	0.0288	0.0364	0.0393
	(0.0007)		(0.0007)		(0.0007)	
Age at the Test Time	0.0147	0.0150	0.0147	0.0153	0.0146	0.0120
	(0.0001)		(0.0001)		(0.0001)	
D=2	-0.9636	-0.9426	-0.9643	-0.9220	-0.9556	-0.9353
	(0.0042)		(0.0043)		(0.0042)	
D=3	-0.9469	-0.9628	-0.9564	-0.9333	-0.9411	-0.9225
	(0.0096)		(0.0084)		(0.0081)	
D=4	-0.8954	-0.9296	-0.8926	-0.8656	-0.9343	-0.8931
	(0.0324)		(0.0512)		(0.0334)	
Residual^2	0.0154	0.0156	0.0154	0.0150	0.0152	0.0156

Table A7

Model Fit: Non-cognitive Skill Measures (CFPS 2010)

	Get on Well with Others		Do you feel Upset often		Feel Everything is Difficult	
	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.0043	0.0309	0.0067	0.0101	0.0067	0.0012
	(0.0030)		(0.0031)		(0.0003)	
Age at the Test Time	0.0028	0.0164	0.0043	0.0001	0.0044	0.0039
	(0.0010)		(0.0001)		(0.0001)	
D=2	-0.1520	-0.1633	-0.2380	-0.2077	-0.2500	-0.7538
	(0.0080)		(0.0089)		(0.0082)	
D=3	-0.1543	-0.4645	-0.2367	-0.2230	-0.2399	-0.2413
	(0.0052)		(0.0090)		(0.0087)	
D=4	-0.1557	-0.3188	-0.2378	-0.2234	-0.2403	-0.2977
	(0.0050)		(0.0060)		(0.0060)	
D=5	-0.1578	-0.1849	-0.2427	-0.2539	-0.2487	-0.5963
	(0.0051)		(0.0057)		(0.0059)	
Residual^2	0.0062	0.0341	0.0091	0.0116	0.0094	0.0236

Table A8

Model Fit: Joint Measures (CFPS)

	Drink 2010		Smoke 2012		Smoke 2014	
	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.4546	0.4382	-0.1206	-0.1288	-0.1083	-0.1137
	(0.1722)		(0.0283)		(0.0278)	
Age at the Test Time	-0.2115	-0.2168	0.0712	0.0734	0.0593	0.0485
	(0.0643)		(0.0106)		(0.0104)	
Whether Smoke	7.3199	3.3107	15.9805	2.0623	15.6562	4.3621
	(1.3610)		(0.2237)		(0.2200)	
Constant	22.3564	22.2506	-0.1406	-0.0491	0.2685	0.4974
	(3.5872)		(0.6081)		(0.6161)	
Residual^2	2685.3307	2685.5114	72.5551	73.1819	70.1309	70.5611

Table A9

Model Fit: Cognitive Skill Measures (CHIP)

	Whether Take College Entrance Exam		Class Performance	
	Data	Model	Data	Model
Years of Education at the Test Time	0.0386	0.0387	-0.0915	-0.0918
	(0.0017)		(0.0049)	
Age at the Test Time	-0.0006	-0.0006	-0.0105	-0.0105
	(0.0005)		(0.0015)	
Constant	-0.2074	-0.2061	3.8114	3.8396
	(0.0306)		(0.0863)	
Residual^2	0.0562	0.0564	0.4464	0.4470

Table A10

Model Fit: Non-cognitive Skill Measures CHIP

	Concentration 2007		Concentration 2008		Whether Confidence in yourself (2007-08)			
	Data	Model	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.0366	0.03337	0.0366	0.03337	0.0367	0.0393	0.0369	0.0371
	(0.0007)		(0.0007)		(0.0007)		(0.0007)	
Age at the Test Time	0.0150	0.0141	0.0150	0.0141	0.0147	0.0145	0.0147	0.0150
	(0.0001)		(0.0001)		(0.0001)		(0.0001)	
D=2	-0.9890	-0.9902	-0.9843	-0.9535	-0.9649	-0.9690	-0.9670	-0.9431
	(0.0048)		(0.0047)		(0.0042)		(0.0042)	
D=3	-0.9921	-0.9050	-0.9939	-0.9797	-0.9362	-0.9391	-0.9451	-0.9532
	(0.0085)		(0.0079)		(0.0165)		(0.0158)	
D=4	-0.9948	-0.9213	-1.0068	-0.9877	-0.9765	-0.8197	-1.0594	-0.9617
	(0.0164)		(0.0151)		(0.0397)		(0.0445)	
Residual^2	0.0158	0.0155	0.0157	0.0160	0.0155	0.0160	0.0155	0.0159

Table A11

Model Fit: Skill Measures CHIP and CFPS

	Concentration			Smoke	
	Data	Model		Data	Model
Years of Education at the Test Time	0.0276	0.0284	Years of Education at the Test Time	-0.1075	-0.1078
	(0.0004)			(0.0396)	
Age at the Test Time	0.0150	0.0154	Age at the Test Time	0.0641	0.0644
	(0.0063)			(0.0141)	
D=2	-0.9502	-0.9392	Whether Smoke	16.0555	16.0557
	(0.0035)			(0.2770)	
D=3	-0.9569	-0.9135	CFPS	2.9963	2.9842
	(0.0063)			(0.2798)	
D=4	-0.9677	-0.9065	Constant	-3.3554	-3.3541
	(0.0104)			(0.8510)	
CFPS	0.1036	0.1029	Residual^2	171.5688	171.5581
	(0.0032)				
Residual^2	0.0251	0.0257			

Table A12

Unobserved Skill Distribution Estimates

Distribution 1		Distribution 2	
Mean (Cognitive)	-0.6950	Probability	0.4120
	(0.0460)		(0.0132)
Mean (Non-Cognitive)	-2.1589	Variance (Cognitive)	0.0620
	(0.0600)		(0.0332)
Variance (Cognitive)	0.0340	Variance (Non-cognitive)	0.1449
	(0.0009)		(0.1556)
Variance (Non-cognitive)	0.7726		
	(0.2046)		